MINIMAX ESTIMATION OF THE MEAN VECTOR OF A SPHERICALLY SYMMETRIC DISTRIBUTION WITH GENERAL QUADRATIC LOSS

by

Ann R. Cohen
University of Minnesota

Technical Report No. 269
June, 1976

ABSTRACT

The problem of estimating the mean θ of a p-dimensional spherically symmetric (s.s.) distribution is investigated. For $p \geq 4$, families of estimators are found whose risks dominate the risk of the best invariant procedure with respect to general quadratic loss, $L(\delta, \theta) = (\delta-\theta)'D(\delta-\theta)$ where D is a $p \times p$ known positive definite matrix. Specifically, if the $p \times 1$ random vector X has a s.s. distribution about θ, then it is proven under quite general conditions that estimators given by $\delta_{a,r,C,D}(X) = (I-(ar(\|X\|^2))D^{-\frac{1}{2}}CD^{\frac{1}{2}}\|X\|^{-2})X$, are minimax estimators with smaller risk than X. For the problem of estimating the mean when n observations $X_1, X_2, \ldots, X_n$ are taken on a p-dimensional s.s. distribution about θ, any spherically symmetric translation invariant estimator, $\delta(X_1, X_2, \ldots, X_n)$, will have a s.s. distribution about θ. It is shown that among the estimators which have these properties are best invariant estimators, sample means and maximum likelihood estimators.

1. Introduction. Charles Stein [18] proved that the best invariant estimator of the mean θ of a multivariate normal distribution with covariance matrix the identity $(MVN(\theta, I))$ is inadmissible in three or more dimensions when the loss is quadratic loss given by

$$(1.1)\qquad \begin{aligned} &L(\delta, \theta) = \|\delta-\theta\|^2 = \sum_{i=1}^{p} (\delta_i-\theta_i)^2 \\ &\text{where } \delta = [\delta_1, \delta_2, \ldots, \delta_p]' \text{ and } \theta = [\theta_1, \theta_2, \ldots, \theta_p]'. \end{aligned}$$

For this case, James and Stein [15], Baranchik [2,3], Alam [1], and Strawderman [19], have all found classes of minimax estimators which are better than the best invariant estimator (which is itself minimax).

In 1966, Brown [11] generalized Stein's inadmissibility result to a larger class of distributions and loss functions. When sampling from a p-dimensional location parameter family, he proved that under mild assumptions on the loss function the best invariant estimator is inadmissible for $p \geq 3$. Investigation of the problem of estimating the mean of a MVN (θ, I) distribution with respect to general quadratic loss

$$(1.2)\qquad \begin{cases} L(\delta, \theta) = (\delta-\theta)'D(\delta-\theta) \\ \text{where } D \text{ is a } p \times p \text{ positive definite matrix} \end{cases}$$

was then taken up. Bhattacharya [7], Bock [8, 9], and Berger [4, 6] were able to find families of estimators which improve on the best invariant procedure in three or more dimensions with respect to the above loss (1.2).

Later on, the problem of finding minimax estimators which are better than the best invariant estimator of the mean of a non-normal location parameter distribution came under investigation. Of interest, was the case when the underlying distribution is spherically symmetric. For certain spherically symmetric non-normal distributions, minimax estimators of the mean were found by Strawderman [20], Berger [5] and Cohen and Strawderman [12]. In this paper, minimax estimators which are better than the best invariant estimator when sampling from any spherically symmetric distribution will be found.

Consider the $p \times 1$ random vector X having a spherically symmetric distribution about θ. The problem of finding estimators which are better than X with respect to quadratic loss (1.1), will be treated in section 2 and the general quadratic loss problem will be taken up in section 3. Specifically, with respect to general quadratic loss, $\delta_{a,r,C,D}(X)$ given by $\delta_{a,r,C,D}(X) = (I-(ar(\|X\|^2))D^{-\frac{1}{2}}CD^{\frac{1}{2}}\|X\|^{-2})X$ where C is a known $p \times p$ positive definite matrix and I is the $p \times p$ identity matrix, will be shown to be better than X provided: (i) $0 < a \leq ((2/p)(\text{trace } CD-2\xi_L)\gamma_L^{-1})/E_0(\|X\|^{-2})$ and ξ_L and γ_L are the maximum eigenvalues of $D^{\frac{1}{2}}CD^{\frac{1}{2}}$ and $D^{\frac{1}{2}}C^2D^{\frac{1}{2}}$, respectively; (ii) $0 < r(\cdot) \leq 1$; (iii) $r(\|X\|^2)$ is nondecreasing; (iv) $r(\|X\|^2)/\|X\|^2$ is non-increasing; and (v) $p \geq 4$. When $p = 3$, there does not exist a strictly positive a for which $\delta_{a,r,C,D}(X)$ is better than X for the general s.s. case. Moreover, this is the largest class of estimators of this form which are better than X for $p \geq 4$.

Of course, the problem would not be complete without considering n observations, $X_1, X_2, \ldots, X_n$, on a p-dimensional s.s. distribution about θ. Section 4 considers the problem of finding estimators which are better than the best invariant estimator and improving on other estimators which are based on n observations. It will be proven that any s.s. translation invariant estimator based on $X_1, X_2, \ldots, X_n$ has a s.s. distribution about θ and best invariant estimators, sample means and maximum likelihood estimators all have these properties. The multiple observation problem is thus reduced to a one observation problem for which there are better estimators than the usual ones.

Note, there is an appendix of useful lemmas (Lemmas A.1-A.7) at the end of the paper which are referred to throughout.

2. Minimax estimators of the location parameter of a p-dimensional $(p \geq 4)$ spherically symmetric distribution with respect to quadratic loss.

Definition 2.1: A $p \times 1$ random vector X is said to have a p-dimensional spherically symmetric (s.s.) distribution about θ if and only if $P(X-\theta)$ has the same distribution as $(X-\theta)$ for every $p \times p$ orthogonal matrix P. The density of X with respect to Lebesque measure will be a function of $||x-\theta||$.

If X has a s.s. distribution about θ, X is the best invariant procedure with respect to quadratic loss (1.1) and it follows from Kiefer [16], that it is a minimax estimator of θ.

Classes of minimax estimators whose risks dominate (are less than or equal to) the risk of X will be found when the loss is sum of squared errors (1.1) and $p \geq 4$.

2.1 James-Stein minimax estimators. When taking one observation, X, on a p-dimensional multivariate normal distribution with mean vector θ and covariance matrix the identity $(X \sim MVN(\theta, I))$, the usual estimator of θ is X. This estimator was improved on for $p \geq 3$ by James and Stein in [15] by estimators of the form

$$(2.1.1) \quad \delta_a(X) = (1-(a/||X||^2))X.$$

For the general case when $X = [X_1, X_2, \ldots, X_p]'$ is one observation on a spherically symmetric distribution about θ, it will be proved that the James-Stein estimator, $\delta_a(X)$ given by (2.1.1), is a "better" estimator of θ than X for $p \geq 4$ when $0 < a \leq ((2(p-2)/p)/E_0(||X||^{-2}))$, where E_0 here and throughout denotes the expectation when $\theta = 0$, and the loss is (1.1). An estimator δ is said to be "better" than X, if the risk of δ dominates the risk of X for all θ with strictly smaller risk than X for some θ. Clearly any estimator which is better than X will be minimax. Thus, the class of estimators, $\delta_a(X)$ for $0 < a \leq (2(p-2)/p)/E_0(||X||^{-2})$ will be a minimax class of estimators for the location parameter of a spherically symmetric distribution.

It is proven in [13] that if the $p \times 1$ random vector X has a s.s. distribution about θ, then the conditional distribution of X given $\|X-\theta\|^2 = R^2$ has a uniform distribution over the surface of the sphere $\{\|X-\theta\|^2 = R^2\}$. Therefore, a natural starting point is to consider one observation, X, on a p-dimensional uniform distribution over the surface of the sphere $\{\|X-\theta\|^2 = R^2\}$ (denoted by $X \sim \mathcal{U}\{\|X-\theta\|^2 = R^2\}$) and show that the risk of $\delta_a(X)$, $R(\delta_a, \theta)$, dominates the risk of X, $R(X, \theta)$ for $p \geq 4$ and $0 < a \leq (2(p-2)/p)/E_0(\|X\|^{-2}) = (2(p-2)/p)R^2$.

For $p \geq 4$ and $0 < a \leq (2(p-2)/p)R^2$, it will be shown that

$$
\begin{aligned}
& R(X, \theta) - R(\delta_a, \theta) \\
(2.1.2) \quad & = E_\theta\|X-\theta\|^2 - E_\theta\|(1-a\|X\|^{-2})X-\theta\|^2 \\
& = aE_\theta[2-2(\theta'X)\|X\|^{-2}-a\|X\|^{-2}]
\end{aligned}
$$

is non-negative for all θ.

If Y is any $p \times 1$ random vector having a s.s. distribution about 0, then the random vector $S = Y\|Y\|^{-1} \sim \mathcal{U}\{\|S\|^2 = 1\}$ (see Dempster [13], page 272). Thus, $RS + \theta = (RY\|Y\|^{-1}) + \theta$ has the same distribution as $X \sim \mathcal{U}\{\|X-\theta\|^2 = R^2\}$. This implies that the distribution of $RY\|Y\|^{-1} + \theta$ is the same as the distribution of $X \sim \mathcal{U}\{\|X-\theta\|^2 = R^2\}$, when Y has a uniform distribution over the entire area of the sphere $\{\|Y\|^2 \leq 1\}$ $(Y \sim \mathcal{U}\{\|Y\|^2 \leq 1\})$. Combine this fact with the expression for the difference in risks given by (2.1.2), and obtain:

$$
\begin{aligned}
& (R(X, \theta) - R(\delta_a, \theta))/a \\
& = E_\theta[2-2(\theta'X)\|X\|^{-2}-a\|X\|^{-2}] \\
& = E[2-(2\theta'(RY\|Y\|^{-1}+\theta)+a)\|RY\|Y\|^{-1}+\theta\|^{-2}]
\end{aligned}
$$

where $Y \sim \mathcal{U}\{\|Y\|^2 \leq 1\}$.

Since Y has the same distribution as PY, where P is any $p \times p$ orthogonal matrix, and there exists an orthogonal transformation Q such that $Q\theta = [\|\theta\|, 0, 0, \ldots, 0]'$, then

$$(2.1.3)\qquad \begin{aligned} &(R(X, \theta) - R(\delta_a, \theta))/a \\ &= E[2-(2RY_1\|\theta\|/\|Y\| + 2\|\theta\|^2+a)(R^2+2RY_1\|\theta\|/\|Y\| + \|\theta\|^2)^{-1}] \end{aligned}$$

where Y_1 is the first coordinate of the random vector Y.

If $Z_1 = Y_1\|Y\|^{-1}$ and $\|\theta\|_1 = \|\theta\|/R$ then (2.1.3) becomes

$$(2.1.4)\qquad \begin{aligned} &(R(X, \theta) - R(\delta_a, \theta))/a \\ &= E[2-(2\|\theta\|_1 Z_1+2\|\theta\|_1^2+(a/R^2))(1+2\|\theta\|_1 Z_1+\|\theta\|_1^2)^{-1}] \\ &= E[(2+2\|\theta\|_1 Z_1-(a/R^2))(1+2\|\theta\|_1 Z_1+\|\theta\|_1^2)^{-1}]. \end{aligned}$$

The density of Z_1, is given by (A.2) in Lemma A.4 as

$$g(z_1) = \begin{cases} M^*(1-z_1^2)^{\frac{p-3}{2}} & \text{when } -1 \le z_1 \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

where $M^* = [\int_0^1 (1-y^2)^{\frac{p-3}{2}} dy]^{-1}/2$. Thus, rewriting the expectations of (2.1.4) in integral form we have

$$(2.1.5)\qquad \begin{aligned} &(R(X, \theta) - R(\delta_a, \theta))/aM^* \\ &= \int_{-1}^{1} (1-z_1^2)^{\frac{p-3}{2}} (2+2\|\theta\|_1 z_1 - aR^{-2})(1+2\|\theta\|_1 z_1+\|\theta\|_1^2)^{-1} dz_1 \\ &= 2\int_0^1 (1-z_1^2)^{\frac{p-3}{2}} ((2-aR^{-2})(1+\|\theta\|_1^2)-4\|\theta\|_1^2 z_1^2)/d(\|\theta\|_1, z_1))dz_1 \\ &= D(a, \|\theta\|_1) \end{aligned}$$

where

$$(2.1.6)\qquad d(\|\theta\|_1, z_1) = (1+\|\theta\|_1^2)^2-4\|\theta\|_1^2 z_1^2 = (1-\|\theta\|_1^2)^2+4\|\theta\|_1(1-z_1^2).$$

The difference in risks will be non-negative for all $\|\theta\|_1$ if and only if $D(a, \|\theta\|_1)$ is non-negative for all $\|\theta\|_1$, or equivalently, if and only if for all $\|\theta\|_1$, $0 \le a \le R^2 b(\|\theta\|_1)$, where

$$(2.1.7)\qquad b(\|\theta\|_1) = \frac{\int_0^1 (1-z_1^2)^{\frac{p-3}{2}} [(2(1+\|\theta\|_1^2)-4\|\theta\|_1^2 z_1^2)/d(\|\theta\|_1, z_1)]dz_1}{\int_0^1 (1-z_1^2)^{\frac{p-3}{2}} [(1+\|\theta\|_1^2)/d(\|\theta\|_1, z_1)]dz_1}.$$

The fact that

$$(2.1.8)\qquad \int_0^1 (1-y^2)^q dy = (2q/(2q+1))\int_0^1 (1-y^2)^{q-1} dy$$

together with the expression for $b(\|\theta\|_1)$ given by (2.1.7) implies

$$b(0) = 2$$

$$b(1) = 2(p-3)/(p-2)$$

and
$$\lim_{\|\theta\|_1 \to \infty} b(\|\theta\|_1) = 2(p-2)/p.$$

Note that when $p = 3$, $b(1) = 0$ implying that $\delta_a(X)$ is not minimax for $a > 0$. For $p \ge 4$, $b(0) > b(1) \ge \lim_{\|\theta\|_1 \to \infty} b(\|\theta\|_1)$. Hence, if it is proven that $\delta_a(X)$ is better than X for $0 < a \le (2(p-2)/p)R^2$, this is the best class of estimators which can be gotten when $X \sim \mathcal{U}\{\|X-\theta\|^2 = R^2\}$ and therefore, the general spherically symmetric result will be the best result for this case.

In order to prove that the risk of $\delta_a(X)$ is smaller than the risk of X for $0 < a \le (2(p-2)/p)R^2$, $D(a, \|\theta\|_1)$ defined by (2.1.5) will be shown to be non-negative for all $\|\theta\|_1$ when $a = (2(p-2)/p)R^2$. Since $D(a, \|\theta\|_1) \ge D((2(p-2)/p)R^2, \|\theta\|_1)$ when $0 < a \le (2(p-2)/p)R^2$, this proves that the risk of

$\delta_a(X)$ dominates the risk of X for all a in that interval.

From (2.1.5) and (2.1.6),

$$D((2(p-2)/p)R^2, \|\theta\|_1) =$$

(2.1.9)
$$(8/p)\int_0^1 (1-z_1^2)^{\frac{p-3}{2}} [((1+\|\theta\|_1^2)-p\|\theta\|_1^2 z_1^2)/d(\|\theta\|_1, z_1)]dz_1.$$

Two cases will be considered, $\|\theta_1\|^2 \leq (p-2)/2$ and $\|\theta_1\|^2 > (p-2)/2$, in order to prove $D((2(p-2)/p)R^2, \|\theta_1\|)$ is non-negative.

Case 1: $\|\theta_1\|^2 \leq (p-2)/2$.

Immediately from the statement of Lemma A.3 the distribution of the random variable Z_1 having the density

$$g_{p-2,1}(z_1) = \begin{cases} ((1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1, z_1))/\int_0^1((1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1, z_1))dz_1 & \\ & \text{when } 0 \leq z_1 \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

has monotone likelihood ratio (MLR) non-decreasing in Z_1 when $\|\theta\|_1 \leq 1$ and MLR non-increasing when $\|\theta\| > 1$.

With respect to the above density, $g_{p-2,1}(z_1)$,

$$D((2(p-2)/p)R^2, \|\theta\|_1)/\int_0^1((1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1, z_1))dz_1$$
$$= D_1((2(p-2)/p)R^2, \|\theta\|_1)$$
$$= (8/p)[(1+\|\theta\|_1^2)-p\|\theta\|_1^2 E_{\|\theta\|_1} Z_1^2]$$
$$= (8/p)[(1-(p-1)\|\theta\|_1^2)+p\|\theta\|_1^2 E_{\|\theta\|_1}(1-Z_1^2)].$$

The MLR properties imply that $E_{\|\theta\|_1}(1-Z_1^2)$ is minimized when $\|\theta\|_1 = 1$ (see Lehmann [17], page 74). By (2.1.8), we easily have $E_1(1-Z_1^2) =$

$\int_0^1 (1-z_1^2)^{\frac{p-3}{2}} dz_1 / \int_0^1 (1-z_1^2)^{\frac{p-5}{2}} dz_1 = (p-3)/(p-2)$. Thus,

$$D_1((2(p-2)/p)R^2,\ \|\theta\|_1)$$

$$= (8/p)[(1-(p-1)\|\theta\|_1^2)+p\|\theta\|_1^2((p-3)/(p-2))]$$

$$= (8/(p(p-2)))[(p-2)-2\|\theta\|_1^2]$$

$$\geq 0$$

when $\|\theta_1\|^2 \leq (p-2)/2$. Hence, for this case, $D_1((2(p-2)/p)R^2, \|\theta\|_1)$ is non-negative and thus $D((2(p-2)/p)R^2, \|\theta\|_1)$ is non-negative.

Case 2: $\|\theta\|_1^2 > (p-2)/2$.

By (2.1.9), and the definition of $d(\|\theta\|_1, z_1)$ given by (2.1.6), it follows immediately that

$$D((2(p-2)/p)R^2,\ \|\theta\|_1)$$

$$(2.1.10)\quad = (2/p)[p\int_0^1 (1-z_1^2)^{\frac{p-3}{2}} dz_1 \ +$$

$$(((4-p)+(4-2p)\|\theta\|_1^2-p\|\theta\|_1^4))\int_0^1 ((1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1,\ z_1))dz_1].$$

Lemma A.2 states that for $\|\theta_1\| > 1$,

$$\int_0^1 ((1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1,\ z_1))dz_1$$

$$= (\int_0^1 (1-z_1^2)^{\frac{p-3}{2}} dz_1/(\|\theta\|_1^2(1+\|\theta\|_1^2)))\sum_{i=0}^{\infty}(-1)^i a_i \|\theta\|_1^{-2i}$$

where

$$a_o = 1 \quad \text{and}$$

$$(2.1.11)\qquad a_i = [(p-2(i+1))/(p+2(i-1))]a_{i-1} \quad \text{for } i = 1, 2, \ldots$$

and $d(\|\theta\|_1, z_1)$ is defined by (2.1.6). Thus, utilizing this expression in (2.1.10), for $p \geq 4$ and $\|\theta\|_1^2 > (p-2)/2$,

$$D((2(p-2)/p)R^2,\ \|\theta\|_1)/((2/p)\|\theta\|_1^{-2}(1+\|\theta\|_1^2)^{-1}\int_0^1(1-z_1^2)^{\frac{p-3}{2}}dz_1)$$

$$= [p\|\theta\|_1^2(1+\|\theta\|_1^2)+((4-p)+(4-2p)\|\theta\|_1^2-p\|\theta\|_1^4)\sum_{i=0}^{\infty}(-1)^i a_i\|\theta\|_1^{-2i}]$$

$$= D_2((2(p-2)/p)R^2,\ \|\theta\|_1).$$

Using the relationship among the a_i's given by (2.1.11), simple calculations lead to

$$D_2((2(p-2)/p)R^2,\ \|\theta\|_1)$$

$$= (4-p)\sum_{i=0}^{\infty}(-1)^i a_i\|\theta\|_1^{-2i}+(2p-4)\sum_{i=0}^{\infty}(-1)^i a_i\|\theta\|_1^{-2i}-p\sum_{i=0}^{\infty}(-1)^i a_i\|\theta\|_1^{-2i}$$

$$+\|\theta\|_1^2(p+(4-2p)+pa_1)+\|\theta\|_1^4(p-p) = \sum_{i=0}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i}$$

where

$$b_i = (4-p)a_i+(2p-4)a_{i+1}-pa_{i+2}$$

$$(2.1.12)\qquad = ([4(p-2)(p-4(i+1)^2)]/[(p+2i)(p+2(i+1))])a_i$$

for $i = 0, 1, 2, \ldots$.

In order to prove $\sum_{i=0}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i}$ is non-negative, which clearly implies $D((2(p-2)/p)R^2,\ \|\theta\|_1)$ is non-negative, a few observations must first be made.

Denote by $K(q)$, the largest integer less than or equal to q, where q is some real number. Thus, $K((\sqrt{p}/2)-1)$ is the largest integer less than or equal to $(\sqrt{p}/2)-1$ and

$$(2.1.13)\quad (p-4(i+1)^2)\begin{cases}\geq 0 & \text{if}\quad 0\leq i\leq K((\sqrt{p}/2)-1)\\ \leq 0 & \text{if}\quad i\geq K((\sqrt{p}/2)-1)+1\end{cases}.$$

Moreover, $K((p-2)/2)$ is the largest integer less than or equal to $(p-2)/2$ and

$$
(2.1.14)\quad
\begin{aligned}
& a_i \geq 0 \quad \text{if} \quad 0 \leq i \leq K((p-2)/2) \\
& (-1)^i a_i \geq 0 \quad \text{if} \quad i \geq K((p-2)/2)+1 \quad \text{and} \quad K((p-2)/2) \quad \text{is an even number} \\
& (-1)^i a_i \leq 0 \quad \text{if} \quad i \geq K((p-2)/2)+1 \quad \text{and} \quad K((p-2)/2) \quad \text{is an odd number.}
\end{aligned}
$$

Combining (2.1.13), (2.1.14) and the definition of b_i given by (2.1.12) clearly

$$
(2.1.15)\quad
\begin{aligned}
& b_i \geq 0 \quad \text{if} \quad 0 \leq i \leq K((\sqrt{p}/2)-1) \\
& b_i \leq 0 \quad \text{if} \quad K((\sqrt{p}/2)-1)+1 \leq i \leq K((p-2)/2) \\
& (-1)^i b_i \leq 0 \quad \text{if} \quad i \geq K((p-2)/2)+1 \quad \text{and} \quad K((p-2)/2) \quad \text{is an even number} \\
& (-1)^i b_i \geq 0 \quad \text{if} \quad i \geq K((p-2)/2)+1 \quad \text{and} \quad K((p-2)/2) \quad \text{is an odd number}
\end{aligned}
$$

If $K((\sqrt{p}/2)-1) = K_1$ is an odd number, rewrite $\sum_{i=0}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i}$ as follows:

$$
(2.1.16)\quad \sum_{i=0}^{K_1-2}(-1)^i b_i\|\theta\|_1^{-2i}+\|\theta\|_1^{-2(K_1-1)}[b_{K_1-1}-b_{K_1}\|\theta\|_1^{-2}+b_{K_1+1}\|\theta\|_1^{-4}]+\sum_{i=K_1+2}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i}
$$

If $K((\sqrt{p}/2)-1) = K_1$ is an even number, rewrite $\sum_{i=0}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i}$ as follows:

$$
(2.1.17)\quad \sum_{i=0}^{K_1-1}(-1)^i b_i\|\theta\|_1^{-2i}+\|\theta\|_1^{-2K_1}[b_{K_1}-b_{K_1+1}\|\theta\|_1^{-2}+b_{K_1+2}\|\theta\|_1^{-4}]+\sum_{i=K_1+3}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i}
$$

The following facts (which can be proven using algebra) will help to clearly show that $\sum_{i=0}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i}$ given previously by (2.1.16) and (2.1.17) is non-negative.

If $0 \leq i \leq K((\sqrt{p}/2)-1)$, then

$$b_i - b_{i+1}\|\theta\|_1^{-2} \geq 0 \quad \text{and if} \quad K((\sqrt{p}/2)-1)+2 \leq i \leq K((p-2)/2)-1, \text{ then}$$

(2.1.18)

$$b_i - b_{i+1}\|\theta\|_1^{-2} \leq 0.$$

If $K_1 = K((\sqrt{p}/2)-1)$ is an odd number then

(2.1.19)

$$b_{K_1-1} - b_{K_1}\|\theta\|_1^{-2} + b_{K_1+1}\|\theta\|_1^{-4} \geq 0.$$

If $K_1 = K((\sqrt{p}/2)-1)$ is an even number then

(2.1.20)

$$b_{K_1} - b_{K_1+1}\|\theta\|_1^{-2} + b_{K_1+2}\|\theta\|_1^{-4} \geq 0$$

If $K_2 = K((p-2)/2)$ then

(2.1.21)

$$\sum_{i=K_2-1}^{\infty} (-1)^i b_i \|\theta\|_1^{-2i} \geq 0.$$

Facts (2.1.18)-(2.1.21) all assume that $\|\theta\|_1 \geq (p-2)/2$ and $p \geq 4$.

Subcase 2.1: $K_1 = K((\sqrt{p}/2)-1)$ is an odd number.

It follows from (2.1.18) that

$$\sum_{i=0}^{K_1-2} (-1)^i b_i \|\theta\|_1^{-2i} \geq 0.$$

Directly by (2.1.19),

$$\|\theta\|_1^{-2(K_1-1)}[b_{K_1-1} - b_{K_1}\|\theta\|_1^{-2} + b_{K_1+1}\|\theta\|_1^{-4}] \geq 0.$$

Lastly, letting $K_2 = K((p-2)/2)$, (2.1.15), (2.1.18) and (2.1.21) imply

$$\sum_{i=K_1+2}^{\infty} (-1)^i b_i \|\theta_1\|^{-2i} = \sum_{i=K_1+2}^{K_2-2} (-1)^i b_i \|\theta_1\|^{-2i} + \sum_{i=K_2-1}^{\infty} (-1)^i b_i \|\theta\|_1^{-2i} \geq 0.$$

It is clear from the expression for $\sum_{i=0}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i}$ given in (2.1.16) that the non-negativity of these three sums implies the non-negativity of $\sum_{i=0}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i}$ for this case.

Subcase 2.2: $K_1 = K((\sqrt{p}/2)-1)$ is an even number.

As in subcase 2.1, from (2.1.18) it follows that

$$\sum_{i=0}^{K_1-1}(-1)^i b_i\|\theta\|_1^{-2i} \geq 0$$

and directly by (2.1.20),

$$\|\theta\|_1^{-2K_1}[b_{K_1}-b_{K_1+1}\|\theta\|_1^{-2}+b_{K_1+2}\|\theta\|_1^{-4}] \geq 0.$$

Once again, letting $K_2 = K((p-2)/2)$, (2.1.15), (2.1.18), and (2.1.21) imply

$$\sum_{i=K_1+3}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i} = \sum_{i=K_1+3}^{K_2-2}(-1)^i b_i\|\theta\|_1^{-2i} + \sum_{i=K_2-1}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i} \geq 0.$$

Thus, clearly by (2.1.17),

$$\sum_{i=0}^{\infty}(-1)^i b_i\|\theta\|_1^{-2i} \geq 0.$$

This, as we noted previously, implies the non-negativity of $D((2(p-2)/p)R^2, \|\theta\|_1)$ which proves that the risk of $\delta_a(X)$ dominates the risk of X for all a in the interval $0 < a \leq (2(p-2)/p)R^2$.

Letting $\theta = 0$ in expression (2.1.3),

$$R(X, 0) - R(\delta_a, 0) = a[2-(a/R^2)].$$

For $0 < a \leq (2(p-2)/p)R^2$, clearly this difference in risks is positive.

It has just been proven that when X is a $p \times 1$ $(p \geq 4)$ random vector distributed uniformly on the sphere $\{\|X-\theta\|^2 = R^2\}$, the estimator of θ given by $\delta_a(X) = (1-a\|X\|^{-2})X$ is better than X with respect to the quadratic loss (1.1), for $0 < a \leq (2(p-2)/p)(1/E_0(\|X\|^{-2}))$.

Now consider the $p \times 1$ random vector X having a spherically symmetric distribution about θ. The conditional distribution of $X \mid \|X-\theta\|^2 = R^2$ (which will be denoted by $X|R$) has a p-dimensional uniform distribution over the sphere $\{\|X-\theta\|^2 = R^2\}$. Thus,

$$(2.1.22) \quad E_0(\|X\|^{-2}) = E[E_0[\|X\|^{-2}|R]] = E(R^{-2}).$$

The following theorem is now ready to be proven.

<u>Theorem 2.1.1:</u> If X is one observation on a p-dimensional $(p \geq 4)$ s.s. distribution about θ and $\delta_a(X)$ is defined by (2.1.1), then with respect to quadratic loss (1.1) $\delta_a(X)$ is better than X for $0 < a \leq$ $(2(p-2)/p)/(E_0(\|X\|^{-2}))$, provided $E_0(\|X\|^{-2})$ is finite.

<u>Proof:</u> With respect to quadratic loss (1.1), when $0 < a \leq$ $(2(p-2)/p)/(E_0(\|X\|^{-2})) = (2(p-2)/p)/E(R^{-2}))$ (as given in 2.1.22))

$$\begin{aligned} & (R(X, \theta) - R(\delta_a, \theta))/a \\ (2.1.23) \quad & = E_\theta[2X'(X-\theta)\|X\|^{-2} - a\|X\|^{-2}] \\ & \geq E_\theta[2X'(X-\theta)\|X\|^{-2} - ((2(p-2)/p)/E(R^{-2}))\|X\|^{-2}]. \end{aligned}$$

Since $X|R$ has a uniform distribution over the sphere $\{\|X-\theta\|^2 = R^2\}$, and for the uniform case $\delta_a(X)$ is better than X for $0 < a \leq (2(p-2)/p)R^2$, then when X has a s.s. distribution about θ

$$\begin{aligned} & E_\theta[2X'(X-\theta)\|X\|^{-2} - (2(p-2)/p)R^2\|X\|^{-2}] \\ (2.1.24) \quad & = E[E_\theta[2X'(X-\theta)\|X\|^{-2} - (2(p-2)/p)R^2\|X\|^{-2}|R]] \\ & \geq 0. \end{aligned}$$

If $E_\theta(R^2\|X\|^{-2}|R)$ is a non-decreasing function of R, then

$$E_\theta\|X\|^{-2}-ER^{-2}E_\theta(R^2\|X\|^{-2})$$

$$= \mathrm{COV}\ (E_\theta(R^2\|X\|^{-2}|R),\ R^{-2})$$

$$\leq 0$$

implying

$$\begin{aligned}(2.1.25)\qquad & E_\theta[2X'(X-\theta)\|X\|^{-2}-(2(p-2)/p)/E_0(R^{-2}))\|X\|^{-2}] \\ & \geq E_\theta[2X'(X-\theta)\|X\|^{-2}-(2(p-2)/p)\ R^2\|X\|^{-2}].\end{aligned}$$

Thus, combining (2.1.23)-(2.1.25), it is clear that the risk of $\delta_a(X)$ dominates the risk of X for all a in the interval $0 < a \leq (2(p-2)/p)/E(R^{-2})$ if $E_\theta(R^2\|X\|^{-2}|R)$ is a non-decreasing function of R. The latter expectation will now be shown to be a non-decreasing function of R.

It is immediate from expressions (2.1.3)-(2.1.5) that

$$(2.1.26)\quad E_\theta(R^2\|X\|^{-2}|R) \propto 2(1+\|\theta\|_1^2)\int_0^1[(1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1,z_1)]dz_1 = I(\|\theta\|_1)$$

where, as in (2.1.6), $d(\|\theta\|_1,\ z_1) = (1+\|\theta\|_1^2)^2-4\|\theta\|_1^2z_1^2$ and $\|\theta\|_1 = \|\theta\|/R$. Moreover, if $I(\|\theta\|_1)$ is non-increasing in $\|\theta\|_1$, then for each fixed $\|\theta\|$, $E_\theta(R^2\|X\|^{-2}|R)$ is non-decreasing in R.

Taking the derivative of $I(\|\theta\|_1)$ with respect to $\|\theta\|_1$,

$$(2.1.27)\quad (d/d\|\theta\|_1)I(\|\theta\|_1) = 2\|\theta\|_1\int_0^1(1-z_1^2)^{\frac{p-3}{2}}((-(1+\|\theta\|_1^2)^2+4z_1^2)/(d(\|\theta\|_1,z_1))^2)dz_1.$$

(By Lebesque's Bounded Convergence Theorem, see Brieman [10], the derivative may be taken inside the integral sign).

It will be shown that this derivative is less than or equal to zero for three cases:

$\|\theta\|_1^2 > 1$; $(2/\sqrt{p-4})-1 \leq \|\theta\|_1^{\,2} \leq 1$, $p \geq 6$; $\|\theta\|_1^2 < (2/\sqrt{p-4})-1$, $p = 6, 7$ and $\|\theta\|_1^2 \leq 1$, $p = 4, 5$.

<u>Case 1:</u> $\|\theta\|_1^2 > 1$.

For this case, $4z_1^2-(1+\|\theta\|_1^2)^2 \leq 4z_1^2-4 \leq 0$ since $0 \leq z_1 \leq 1$. Thus, clearly $(d/d\|\theta\|_1)I(\|\theta\|_1)$ given by (2.1.26) is non-positive.

<u>Case 2:</u> $(2/\sqrt{p-4})-1 \leq \|\theta\|_1^2 \leq 1$, $p \geq 6$.

With respect to the density $g_{p-2,2}(z_1)$, where

$$g_{p-2,2}(z_1) = \begin{cases} [(1-z_1^2)^{\frac{p-3}{2}}/(d(\|\theta\|_1, z_1))^2]/[\int_0^1((1-z_1^2)^{\frac{p-3}{2}}/(d(\|\theta\|_1,z_1))^2)dz_1] & \text{for } 0 \leq z_1 \leq 1 \\ 0 & \text{elsewhere,} \end{cases}$$

$$(d/d\|\theta\|_1)I(\|\theta\|_1)/[\int_0^1((1-z_1^2)^{\frac{p-3}{2}}/((d(\|\theta\|_1, z_1))^2dz_1]$$

$$= 2\|\theta\|_1[-(1+\|\theta\|_1^2)^2+E_{\|\theta\|_1}4Z_1^2].$$

As stated in Lemma A.3, $g_{p-2,2}(z_1)$ has MLR non-decreasing in z_1 when $\|\theta\|_1 \leq 1$. Thus, $E_{\|\theta\|_1}Z_1^2$ is maximized when $\|\theta\|_1 = 1$ and with the aid of (2.1.8), it is seen that that maximum value is $1/(p-4)$. Thus,

$$2\|\theta\|_1[-(1+\|\theta\|_1^2)^2+E_{\|\theta\|_1}4Z_1^2]$$

$$\leq (2\|\theta\|_1/(p-4))[4-(p-4)(1+\|\theta\|_1^2)^2]$$

$$\leq 0.$$

So, for this case, $(d/d\|\theta\|_1)I(\|\theta\|_1)$ is non-positive.

<u>Case 3:</u> $\|\theta_1\|^2 < (2/\sqrt{p-4})-1$, $p = 6,7$ and $\|\theta\|_1^2 \leq 1$, $p = 4,5$.

Directly by Lemma A.2,

$$I(\|\theta\|_1) = 2\left(\int_0^1 (1-z_1^2)^{\frac{p-3}{2}} dz_1\right) \sum_{i=0}^{\infty}(-1)^i a_i \|\theta\|_1^{2i}$$

where $a_0 = 1$ and $a_i = [(p-2(i+1))/(p+2(i-1))]\ a_{i-1}$ for $i = 1, 2, \ldots$ From this expression, the derivative of $I(\|\theta\|_1)$ is

$$(d/d\|\theta\|_1)I(\|\theta\|_1) \propto (2/\|\theta\|_1) \sum_{i=1}^{\infty}(-1)^i i a_i \|\theta\|_1^{2i}.$$

Using the relationship among the a_i's given above and those given in (2.1.14), it is simple to show that

(2.1.28) $(-1)^i[ia_i-(i+1)a_{i+1}] \geq 0$ when $p = 6, 7$ and $i \geq 2$.

When $p = 6,7$, $K((p-2)/2) = 2$ and thus, by (2.1.14), a_1 and a_2 are non-negative. Since $\|\theta\|_1^2 < (2/\sqrt{p-4})-1$, this together with (2.1.28) implies

$$\sum_{i=1}^{\infty}(-1)^i i a_i \|\theta\|_1^{2i} = -a_1\|\theta\|_1^2 + \sum_{i=2}^{\infty}(-1)^i i a_i \|\theta\|_1^{2i}$$

$$\leq -a_1\|\theta\|_1^2 + \sum_{i=2}^{\infty}(-1)^2 2a_2\|\theta\|_1^{2i} = -a_1\|\theta\|_1^2 + 2a_2[\|\theta\|_1^2/(1-\|\theta\|_1^2)]$$

$$= a_1(\|\theta\|_1^2(1-\|\theta\|_1^2)^{-1}(p+2)^{-1})[(p+2)\|\theta\|_1^2-8] \leq 0.$$

When $p = 4,5$, $K((p-2)/2) = 1$ and thus, by (2.1.14), $(-1)^i a_i$ is less than or equal to zero for $i \geq 2$ and a_1 is non-negative. Thus,

$$\sum_{i=1}^{\infty}(-1)^i i\ a_i\|\theta\|_1^{2i} = -a_1 + \sum_{i=2}^{\infty}(-1)^i i a_i \|\theta\|_1^{2i} \leq 0.$$

Hence, for this case,

$$(d/d\|\theta\|_1)I(\|\theta\|_1) \leq 0.$$

Thus, the risk of $\delta_a(X)$ dominates the risk of X for all a between 0 and $(2(p-2)/p)/ER^{-2}$. Moreover, from (2.1.23) it is immediate that

$$(R(X, 0) - R(\delta_a, 0))/a$$

$$\geq 2-2(p-2)/p = 4/p > 0,$$

completing the proof of this theorem.

<u>2.2 A larger class of estimators which are better than the best invariant procedure.</u> In [2], Baranchik found classes of estimators which include the James-Stein class and improve on the best invariant estimator of the mean of a p-dimensional ($p \geq 3$) multivariate normal distribution when the loss is sum of squared errors (1.1). When the $p \times 1$ random vector X has a $MVN(\theta, I)$ distribution, the Baranchik estimators are of the form

$$(2.2.1) \quad \delta_{a,r}(X) = (1-a(r(\|X\|^2)/\|X\|^2))X$$

where $r(\|X\|^2)$ is a non-decreasing function between 0 and 1, inclusive. When $r(\cdot)$ is identically 1, this is just $\delta_a(X)$ given by (2.1.1).

If $X = [X_1, X_2, \ldots, X_p]'$ has a s.s. distribution about θ and $p \geq 4$, under certain conditions estimators of the form given by (2.2.1) will be better than X and thus minimax. The conditions are stated in the following theorem.

<u>Theorem 2.2.1:</u> If X has a p-dimensional ($p \geq 4$) spherically symmetric distribution about θ, then with respect to quadratic loss (1.1), the risk of $\delta_{a,r}(X)$, where $\delta_{a,r}(X)$ is defined above by (2.2.1), is less than or equal to the risk of X for all θ with strict inequality for some θ provided:

(1) $0 < r(\cdot) \leq 1$

(2) $r(\|X\|^2)$ is non-decreasing

(3) $r(\|X\|^2)/\|X\|^2$ is non-increasing

and

(4) $0 < a \leq (2(p-2)/p)/E_0(\|X\|^{-2})$.

<u>Proof:</u> It is natural to begin by proving this theorem for $X \sim \mathcal{U}\{\|X-\theta\|^2 = R^2\}$.

Again, it will be shown that the difference in risks, $R(X, \theta) - R(\delta_{a,r}, \theta)$ is non-negative, where

$$(R(X, \theta) - R(\delta_{a,r}, \theta))/a$$

$$= E_\theta[r(||X||^2)(2-2(X'\theta)||X||^{-2}-ar(||X||^2)||X||^{-2})].$$

Since $0 < r(\cdot) \leq 1$ and $||X-\theta||^2 = ||X||^2 - 2X'\theta + ||\theta||^2 = R^2$, immediately

$$(R(X, \theta) - R(\delta_{a,r}, \theta))/a$$

$$\geq E_\theta[r(||X||^2)(2-(2X'\theta)||X||^{-2}-a||X||^{-2})]$$

(2.2.2)

$$= E_\theta[r(||X||^2)(2+(R^2-||\theta||^2-||X||^2-a)||X||^{-2}]$$

$$= E_\theta[r(||X||^2)(1+(R^2-||\theta||^2-a)||X||^{-2}].$$

For $0 < a \leq (2(p-2)/p)/E_0(||X||^{-2}) = (2(p-2)/p)R^2$, using the expressions given in (2.2.2),

$$(R(X, \theta) - R(\delta_{a,r}, \theta))/a$$

$$\geq E_\theta[r(||X||^2)(1-(1/p)(p||\theta||^2+(p-4)R^2)||X||^{-2})]$$

$$\geq E_\theta[r(||X||^2)]E_\theta[1-(1/p)(p||\theta||^2+(p-4)R^2)||X||^{-2}].$$

The last statement follows immediately since $r(||X||^2)$ is non-decreasing (property (2)), $||X||^{-2}$ is non-increasing and for $p \geq 4$, $(p||\theta||^2+(p-4)R^2)$ is always non-negative.

Now, $E_\theta[1-(1/p)(p||\theta||^2+(p-4)R^2)||X||^{-2}]$ is just $(R(X, \theta) - R(\delta_a, \theta))/a$ when $a = (2(p-2)/p)R^2$. This was proven to be non-negative in the previous section, thus, the theorem is true for the uniform case.

For the general case when X has a s.s. distribution about θ, recall that the conditional distribution of X given the $||X-\theta||^2 = R^2$ (once again denoted by $X|R$)

has a uniform distribution on the surface of the sphere $\{\|X-\theta\|^2 = R^2\}$. Thus, for the general case $(2(p-2)/p)/E_0(\|X\|^{-2}) = (2(p-2)/p)/ER^{-2}$ and the theorem is true if

$$(R(X,\ \theta) - R(\delta_{a,r},\ \theta))/a$$

(2.2.3) $$\geq E_\theta[r(\|X\|^2)(2-2(X'\theta)\|X\|^{-2}-a\|X\|^{-2}]$$

$$\geq E_\theta[r(\|X\|^2)(2-2(X'\theta)\|X\|^{-2}-((2(p-2)/p)/E(R^{-2}))\|X\|^{-2}]$$

is non-negative.

Using the result for the uniform case, for the general s.s. case,

$$E_\theta[r(\|X\|^2)(2-2(X'\theta)\|X\|^{-2}-(2(p-2)/p)R^2\|X\|^{-2}]$$

(2.2.4) $$= E[E_\theta[r(\|X\|^2)(2-2(X'\theta)\|X\|^{-2}-(2(p-2)/p)R^2\|X\|^{-2}|R]]$$

$$\geq 0.$$

If $E_\theta[(r(\|X\|^2)R^2\|X\|^{-2})|R]$ is non-decreasing in R, then

$$E_\theta[r(\|X\|^2)\|X\|^{-2}]$$

$$= E[(R^{-2})E_\theta[(r(\|X\|^2)R^2\|X\|^{-2})|R]]$$

$$\leq E(R^{-2})E_\theta(r(\|X\|^2)R^2\|X\|^{-2})$$

implying together with (2.2.3) and (2.2.4)

$$(R(X,\ \theta) - R(\delta_{a,r},\ \theta))/a$$

$$\geq E_\theta[r(\|X\|^2)(2-2(X'\theta)\|X\|^{-2}-((2(p-2)/p)/E(R^{-2}))\|X\|^{-2}]$$

$$\geq E_\theta[r(\|X\|^2)(2-2(X'\theta)\|X\|^{-2}-((2(p-2)/p)R^2\|X\|^{-2}]$$

$$\geq 0.$$

Hence the proof of the theorem is complete once $E_\theta[(r(\|X\|^2)R^2\|X\|^{-2})|R]$ is shown to be non-decreasing in R (this also implies $R(X, 0) - R(\delta_{a,r}, 0)$ is clearly positive). This will now be proven.

Let the $p \times 1$ random vector Y be distributed uniformly on the surface of the sphere $\{\|Y-\theta\|^2 = R^2\}$. Then, $E_\theta[(r(\|X\|^2)R^2\|X\|^{-2})|R] = E_\theta[r(\|Y\|^2)R^2\|Y\|^{-2}]$ $= E[r(\|RZ\ \|Z\|^{-1}+\theta\|^2)R^2\|RZ\|Z\|^{-1}+\theta\|^{-2}]$ where $Z = [Z_1, Z_2, \ldots, Z_p]'$ has a uniform distribution over the entire sphere $\{\|Z\|^2 \leq 1\}$. If we let $S_1 = Z_1\|Z\|^{-1}$ then,

$$E_\theta[(r(\|X\|^2)R^2\|X\|^{-2})|R]$$

$$= E[r\,(R^2+2R\|\theta\|Z_1\|Z\|^{-1}+\|\theta\|^2)R^2(R^2+2R\|\theta\|Z_1\|Z\ \|^{-1}+\|\theta\|^2)^{-1}] \tag{2.2.5}$$

$$= E[r(R^2+2R\|\theta\|S_1+\|\theta\|^2)R^2(R^2+2R\|\theta\|S_1+\|\theta\|^2)^{-1}].$$

In order to prove that $E_\theta[(r(\|X\|^2)R^2\|X\|^{-2})|R]$ is non-decreasing the cases when the $\|\theta\|$ is greater than R and the $\|\theta\|$ is less than or equal to R will be considered separately.

<u>Case 1:</u> $\|\theta\| > R$.

For this case, since S_1 falls between plus and minus 1, clearly, $(R^2+2R\|\theta\|S_1+\|\theta\|^2)/R^2$ is non-increasing in R for fixed $\|\theta\|$. However, properties (2) and (3) of this theorem state that $r(\|X\|^2)$ is non-decreasing and $r(\|X\|^2)/\|X\|^2$ is non-increasing, thus, $r(R^2(R^2+2R\|\theta\|S_1+\|\theta\|^2)/R^2))$ divided by $((R^2+2R\|\theta\|S_1+\|\theta\|^2)/R^2)$ is non-decreasing in R. Utilizing this fact, clearly, $E_\theta[r(\|X\|^2)R^2\|X\|^{-2}|R]$ given above by (2.2.5) is non-decreasing in R for fixed θ, which proves the theorem for this case.

<u>Case 2:</u> $\|\theta\| \leq R$.

For this second case, $R^2+2\|\theta\|RS_1+\|\theta\|^2$ is non-decreasing in R for fixed $\|\theta\|$ and so property (2) of this theorem implies that $r(R^2+2\|\theta\|RS_1+\|\theta\|^2)$ is non-decreasing in R. Since $r(R^2+2R\|\theta\|S_1+\|\theta\|^2)$ is non-decreasing in R,

$$(d/dR)r(R^2+2\|R\|\,\|\theta\|S_1+\|\theta\|^2)$$

$$(2.2.6) \qquad = (2R+2\|\theta\|S_1)r'(R^2+2\|\theta\|RS_1+\|\theta\|^2)$$

$$\geq 0.$$

Moreover, in the proof of Theorem 2.1.1, it was shown that $E_\theta(R^2\|X\|^{-2}|R)$ is a non-decreasing function of R. Thus,

$$(d/dR)E_\theta(R^2\|X\|^{-2}|R)$$

$$= (d/dR)E[R^2/(R^2+2R\|\theta\|S_1+\|\theta\|^2)]$$

$$(2.2.7) \qquad = E[2R\|\theta\|(RS_1+\|\theta\|)/(R^2+2R\|\theta\|S_1+\|\theta\|^2)^2]$$

$$\geq 0.$$

Taking the derivative with respect to R of $E_\theta[R^2r(\|X\|^2)\|X\|^{-2}|R]$

$$(d/dR)E_\theta[R^2r(\|X\|^2)\|X\|^{-2}|R]$$

$$= (d/dR)E[r(R^2+2R\|\theta\|S_1+\|\theta\|^2)R^2/(R^2+2R\|\theta\|S_1+\|\theta\|^2)]$$

$$(2.2.8) \qquad = E[r(R^2+2R\|\theta\|S_1+\|\theta\|^2)2R\|\theta\|(RS_1+\|\theta\|)/(R^2+2R\|\theta\|S_1+\|\theta\|^2)^2]$$

$$+ E[(2R+2\|\theta\|S_1)r'(R^2+2R\|\theta\|S_1+\|\theta\|^2)R^2/(R^2+2R\|\theta\|S_1+\|\theta\|^2)].$$

Combining (2.2.6) with (2.2.8),

$$(d/dR)E_\theta[R^2r(\|X\|^2)\|X\|^{-2}|R]$$

$$(2.2.9) \qquad \geq E[r(R^2+2R\|\theta\|S_1+\|\theta\|^2)2R\|\theta\|(RS_1+\|\theta\|)/(R^2+2R\|\theta\|S_1+\|\theta\|^2)^2].$$

Property (2) states $r(\|X\|^2)$ is non-decreasing. This together with (2.2.7) and (2.2.9) implies

$$(d/dR)E_\theta[R^2 r(\|X\|^2)\|X\|^{-2}|R]$$

$$\geq r(R^2-\|\theta\|^2)E[2R\|\theta\|(RS_1+\|\theta\|)/(R^2+2R\|\theta\|S_1+\|\theta\|^2)^2]$$

$$\geq 0.$$

For this case, $E_\theta[R^2 r(\|X\|^2)\|X\|^{-2}|R]$ is non-decreasing in R.

The proof of the theorem is now complete.

3. Minimax estimators of the location parameter of a p-dimensional $(p \geq 4)$ spherically symmetric distribution with respect to general quadratic loss.

For one $p \times 1$ observation, X, on a s.s. distribution about θ, minimax estimators which are better than X were found in the previous section for quadratic loss (1.1). In this section, those results will be explicitly extended for general quadratic loss given, as in (1.2), by

$$L(\delta, \theta) = (\delta-\theta)'D(\delta-\theta)$$

where D is a known $p \times p$ positive definite matrix.

3.1. James-Stein and Bock minimax estimators for general quadratic loss.

When X has a p-variate $(p \geq 3)$ $MVN(\theta, I)$ distribution, with respect to general quadratic loss (1.2), X is minimax and inadmissible. In [9], Bock found values of a for which the James-Stein estimator $\delta_a(X)$ given by (2.1.1) is a better estimator of θ than X. Later on, in [8], Bock found a larger class of estimators which are better than X and are of the following form:

$$(3.1.1) \qquad \delta_{a,C,D}(X) = (I-(a(D^{-\frac{1}{2}}CD^{\frac{1}{2}})/\|X\|^2))X$$

where C is a known $p \times p$ positive definite matrix, I is the $p \times p$ identity matrix, and $D^{\frac{1}{2}}D^{\frac{1}{2}} = D$ is the known $p \times p$ matrix in the loss function. Of course, when C equals the identity, $\delta_{a,C,D}(X)$ just becomes the James-Stein estimator $\delta_a(X)$.

For X one observation on a p-dimensional $(p \geq 4)$ s.s. distribution about θ, estimators of θ of the form (3.1.1) given above will be shown to be better than X for certain values of a. Since X itself is minimax, these estimators will certainly be minimax. In addition, when C and D are both the identity, this class of estimators will coincide with the class given in section 2.1, and thus will be the largest class of estimators of this form for the general case.

The following theorem details all the conditions under which $\delta_{a,C,D}(X)$ is better than X.

Theorem 3.1.1: If $X = [X_1, X_2, \ldots, X_p]$ is a single observation on a spherically symmetric distribution about θ and $\delta_{a,C,D}(X)$ is given by (3.1.1) then, with respect to general quadratic loss (1.2), the risk of $\delta_{a,C,D}(X)$ dominates (is less than or equal to) the risk of X when $p \geq 4$ provided

(1) trace $CD = \text{tr}CD > 2\xi_L = 2$(maximum eigenvalue of $D^{\frac{1}{2}}CD^{\frac{1}{2}}$)

and (2) $0 < a \leq ((2/p)(\text{tr}CD - 2\xi_L)\gamma_L^{-1})/E_0(\|X\|^{-2})$

where γ_L = maximum eigenvalue of $D^{\frac{1}{2}}C^2D^{\frac{1}{2}}$.

Moreover, under these conditions the risk of $\delta_{a,C,D}(X)$ is strictly less than the risk of X when θ is the zero vector, and so, $\delta_{a,C,D}(X)$ is a better estimator of θ than X.

Proof: First, the theorem will be proven for $X \sim \mathcal{U}\{\|X-\theta\|^2 = R^2\}$. The more general result will then fall out immediately by once again using the fact that when X has a s.s. distribution about θ, the conditional distribution of X given $\|X-\theta\|^2 = R^2$ (denoted by $X|R$) is uniform on the surface of the sphere $\{\|X-\theta\|^2 = R^2\}$.

For $X \sim \mathcal{U}\{\|X-\theta\|^2 = R^2\}$, the difference in risks with respect to general quadratic loss, between X and $\delta_{a,C,D}(X)$ is

$$\begin{aligned}
&(RD(X, \theta) - RD(\delta_{a,C,D}, \theta)) \\
(3.1.2) \quad &= E_\theta[(X-\theta)'D(X-\theta)] - E_\theta[(X-\theta-(a\|X\|^{-2})D^{-\frac{1}{2}}CD^{\frac{1}{2}}X)'D(X-\theta-(a\|X\|^{-2})D^{-\frac{1}{2}}CD^{\frac{1}{2}}X)] \\
&= aE_\theta[2X'D^{\frac{1}{2}}CD^{\frac{1}{2}}(X-\theta)\|X\|^{-2} - a(X'D^{\frac{1}{2}}C^2D^{\frac{1}{2}}X)\|X\|^{-4}]
\end{aligned}$$

Throughout, as in (3.1.2), the risk with respect to general quadratic loss will be denoted by RD. Now, if $Y \sim \mathcal{U}\{\|Y\|^2 \leq 1\}$, then $RY\|Y\|^{-1} + \theta$ has the same distribution as X, and so (3.1.2) becomes

$$(RD(X,\ \theta)-RD(\delta_{a,C,D},\ \theta))$$
$$= aE[2(RY\|Y\|^{-1}+\theta)'D^{\frac{1}{2}}CD^{\frac{1}{2}}\ RY\|Y\|^{-1}\|RY\|Y\|^{-1}+\theta\|^{-2}$$
$$- a(RY\|Y\|^{-1}+\theta)'D^{\frac{1}{2}}C^2D^{\frac{1}{2}}(RY\|Y\|^{-1}+\theta)\|RY\|Y\|^{-1}+\theta\|^{-4}]$$
$$= aE[2(Z+\theta_R)'D^{\frac{1}{2}}CD^{\frac{1}{2}}Z\|Z+\theta_R\|^{-2}-aR^{-2}(Z+\theta_R)'D^{\frac{1}{2}}C^2D^{\frac{1}{2}}(Z+\theta_R)\|Z+\theta_R\|^{-4}]$$

where $Y\|Y\|^{-1} = [Z_1,\ Z_2,\ \ldots,\ Z_p]' = Z$ and $\theta_R = \theta/R$.

There exists a $p \times p$ orthogonal matrix P such that $P'D^{\frac{1}{2}}CD^{\frac{1}{2}}P = D(\xi_i)$ where $\xi_1,\ \xi_2,\ \ldots,\ \xi_p$ are the eigenvalues of $D^{\frac{1}{2}}CD^{\frac{1}{2}}$ and $D(\xi_i)$ is the diagonal matrix whose diagonal elements are the eigenvalues $\xi_1,\ \xi_2,\ \ldots,\ \xi_p$. Similarly, there exists a $p \times p$ orthogonal matrix Q such that $Q'D^{\frac{1}{2}}C^2D^{\frac{1}{2}}Q = D(\gamma_i)$ where $\gamma_1,\ \gamma_2,\ \ldots,\ \gamma_p$ are the eigenvalues of $D^{\frac{1}{2}}C^2D^{\frac{1}{2}}$. Moreover, let $\theta^* = P'\theta_R = [\theta_1^*,\ \theta_2^*,\ \ldots,\ \theta_p^*]$ and $\theta^{**} = Q'\theta_R = [\theta_1^{**},\ \theta_2^{**},\ \ldots,\ \theta_p^{**}]'$. Of course $P'Z$ and $Q'Z$ have the same distribution as Z, so

$$(RD(X,\ \theta)-RD(\delta_{a,C,D},\ \theta))/a$$
$$(3.1.3)\qquad = \sum_{i=1}^{p}\xi_i E[2(Z_i^2+\theta_i^*Z_i)\ \|Z+\theta^*\|^{-2}]$$
$$-aR^{-2}\sum_{i=1}^{p}\gamma_i E[(Z_i+\theta_i^{**})^2\ \|Z+\theta^{**}\|^{-4}].$$

Clearly, $\|\theta_R\| = \|\theta^*\| = \|\theta^{**}\|$ and if $\|T\|^2 = \sum_{i=2}^{p} Z_i^2$ then $1 = \|Z\|^2 = Z_1^2 + \|T\|^2$.

From these observations and expressions for the expectations in (3.1.3) - given generally by (A.3) and (A.4) of Lemma (A.5), (3.1.3) multiplied by (p-1) becomes

$$(RD(X,\ \theta)-RD(\delta_{a,C,D},\ \theta))((p-1)/a)$$

$$= \sum_{i=1}^{p} \xi_i \theta_i^{*2} \|\theta_R\|^{-2} E[(2(p-1)(Z_1^2+\|\theta_R\|Z_1)-2\|T\|^2)((Z_1+\|\theta_R\|)^2+\|T\|^2)^{-1}]$$

$$(3.1.4) \qquad +2\sum_{i=1}^{p} \xi_i E[\|T\|^2((Z_1+\|\theta_R\|)^2+\|T\|^2)^{-1}]$$

$$-aR^{-2}[\sum_{i=1}^{p} \gamma_i (\theta_i^{**})^2 \|\theta_R\|^{-2} E[((p-1)(Z_1+\|\theta_R\|)^2-\|T\|^2)((Z_1+\|\theta_R\|)^2+\|T\|^2)^{-2}]$$

$$+\sum_{i=1}^{p} \gamma_i E[\|T\|^2((Z_1+\|\theta_R\|)^2+\|T\|^2)^{-2}]]$$

For consistency with section 2, denote hereon, $\|\theta_R\| = \|\theta\|R^{-1}$ by $\|\theta\|_1$. Lemma A.6 states that $E[(2(p-1)(Z_1^2+\|\theta\|_1 Z_1)-2\|T\|^2)((Z_1+\|\theta\|_1)^2+\|T\|^2)^{-1}]$ is less than or equal to zero and Lemma A.7 that $E[((p-1)(Z_1+\|\theta\|_1)^2-\|T\|^2)((Z_1+\|\theta\|_1)^2+\|T\|^2)^{-2}]$ is non-negative. Moreover, $0 < a \le (2/p)(\mathrm{tr}CD-2\xi_L)\gamma_L^{-1}R^2$ and $(\mathrm{tr}C^2D)\gamma_L^{-1}$ is less than or equal to p. It now follows from (3.1.4) that

$$(RD(X,\ \theta)-RD(\delta_{a,C,D},\ \theta))((p-1)/a)$$

$$\ge \xi_L E[(2(p-1)(Z_1^2+\|\theta\|_1 Z_1)-2\|T\|^2)((Z_1+\|\theta\|_1)^2+\|T\|^2)^{-1}]$$

$$+2(\mathrm{tr}CD)E[\|T\|^2((Z_1+\|\theta\|_1)^2+\|T\|^2)^{-1}]$$

$$-(2/p)(\mathrm{tr}CD-2\xi_L)\gamma_L^{-1}[\gamma_L E[((p-1)(Z_1+\|\theta\|_1)^2-\|T\|^2)((Z_1+\|\theta\|_1)^2+\|T\|^2)^{-2}]$$

$$(3.1.5) \qquad +(\mathrm{tr}C^2D)E[\|T\|^2((Z_1+\|\theta\|_1)^2+\|T\|^2)^{-2}]]$$

$$\ge \xi_L[E[((2(p-1)(Z_1^2+\|\theta\|_1 Z_1)-2\|T\|^2)((Z_1+\|\theta\|_1)^2+\|T\|^2)^{-1}]$$

$$+2qE[\|T\|^2((Z_1+\|\theta\|_1)^2+\|T\|^2)^{-1}]$$

$$-(2/p)(q-2)[(p-1)E((Z_1+\|\theta\|_1)^2+\|T\|^2)^{-1}]]$$

$$= \xi_L \Delta_q'$$

where $q = (\mathrm{trCD}/\xi_L)$. Using the expression for the density of Z_1 given in Lemma A.4 and since $\|T\|^2 = 1-Z_1^2$,

$$
(3.1.6)\quad
\begin{aligned}
(d/dq)\Delta q &= (2/p)E[(1-pZ_1^2)(1+2\|\theta\|_1 Z_1+\|\theta\|_1^2)^{-1}] \\
&\propto (1+\|\theta\|_1)^2\int_0^1 (1-z_1^2)^{\frac{p-3}{2}}(1-pz_1^2)/((1+\|\theta\|_1^2)^2-4\|\theta\|_1^2 z_1^2)dz_1.
\end{aligned}
$$

With respect to the density

$$
g_{p-2,1}(z_1) = \begin{cases} (1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1, z_1)/\int_0^1((1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1, z_1))dz_1 & \text{for } 0 \leq z_1 \leq 1 \\ 0 & \text{elsewhere} \end{cases}
$$

where $d(\|\theta\|_1, z_1) = (1+\|\theta\|_1^2)-4\|\theta\|_1^2 z_1^2$, the derivative given in (3.1.6) divided by $\int_0^1((1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1, z_1))dz_1$ is $(1+\|\theta\|_1^2)E_{\|\theta\|_1}(1-pz_1^2)$. Moreover, by Lemma A.3, the distribution of Z_1 has MLR non-decreasing in Z_1 when $\|\theta\|_1 \leq 1$ and non-increasing in Z_1 when $\|\theta\|_1 > 1$, thus,

$$
E_{\|\theta\|_1} Z_1^2 \geq E_0 Z_1^2 = \lim_{\|\theta\|_1\to\infty} E_{\|\theta\|_1} Z_1^2 = (1/p)
$$

and so $E_{\|\theta\|_1}(1-pZ_1^2) \leq 0$ implying $(d/dq)\Delta q$ is less than or equal to zero, and thus Δ_q is a decreasing function of q, for fixed $\|\theta\|_1$. Hence, since q is between 2 and p,

$$
\begin{aligned}
\Delta_q &\geq \Delta_p \\
&= \frac{2(p-1)}{p} E[(2+p\|\theta\|_1 Z_1)(1+2\|\theta\|_1 Z_1+\|\theta\|_1^2)^{-1}] \\
&= (p-1)(\text{Difference in risks given by (2.1.4) when } a=(2(p-2)/p)R^2) \\
&\geq 0 \quad \text{(as proven in section 2.1).}
\end{aligned}
$$

Thus, since ξ_L is positive, $(RD(X, \theta)-RD(\delta_{a,C,D}, \theta))$ is non-negative for $0 < a \leq 2(\text{tr } CD-2\xi_L)\gamma_L^{-1}R^2$, when $X \sim \mathcal{U}\{\|X-\theta\|=R^2\}$.

When $X \sim$ s.s. about θ, since $X|R \sim \mathcal{U}\{\|X-\theta\|^2=R^2\}$ and $E_0(\|X\|^{-2}) = E(R^{-2})$, clearly by the work just done, for $0 < a \leq 2(\text{tr } CD-2\xi_L)\gamma_L^{-1}/E(R^{-2})$,

$$(RD(X, \theta)-RD(\delta_{a,C,D}, \theta))((p-1)/a)$$

$$\begin{aligned} &\geq \xi_L[E[(2(p-1)(Z_1^2+\|\theta\|R^{-1}Z_1)-2\|T\|^2)((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)^{-1}] \\ &\quad +2qE[\|T\|^2((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)^{-1}] \\ &\quad -(2/p)(q-2)(E(R^{-2}))^{-1}E[(p-1)R^{-2}((Z_1+\|\theta\|R^{-1})^2 + \|T\|^2)^{-1}]] \end{aligned} \tag{3.1.7}$$

where $q = \text{tr } CD/\xi_L$ and $Z = [Z_1, Z_2, \ldots, Z_p] \sim \mathcal{U}\{\|Z\|^2=1\}$. Now $E_\theta[R^2\|X\|^{-2}|R]$, as proven in Theorem 2.1.1, is a non-decreasing function of R and

$$E_\theta[R^2\|X\|^{-2}|R]$$

$$= E[((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)^{-1}|R]$$

implying $E[R^{-2}((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)^{-1}]$

$$\leq E(R^{-2})E[((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)^{-1}].$$

This, together with (3.1.7), leads to

$$(RD(X, \theta)-RD(\delta_{a,C,D}, \theta))((p-1)/a)$$

$$\geq \xi_L E(\Delta_q)$$

$$\geq 0$$

since Δ_q was just proven to be non-negative.

Clearly, when $\theta = 0$,

$$(R(X,\ 0)-R(\delta_{a,C,D},\ 0))((p-1)/a)$$

$$\geq \xi_L E(\Delta_p)$$

$$= \xi_L(4(p-1)/p) > 0.$$

Thus, the risk of $\delta_{a,C,D}(X)$ dominates the risk of X for all θ and is strictly smaller when θ is 0. The proof of this theorem is now complete.

3.2. A larger class of minimax estimators for the mean of a p-dimensional ($p \geq 4$) s.s. distribution. For X one observation on a p-dimensional ($p \geq 4$) s.s. distribution about θ, consider the estimator

$$(3.2.1)\qquad \delta_{a,r,C,D}(X) = (I-(ar(\|X\|^2)D^{-\frac{1}{2}}CD^{\frac{1}{2}}/\|X\|^2))X$$

where C is a known $p \times p$ positive definite matrix, I is the $p \times p$ identity matrix, and D is the known $p \times p$ matrix in the loss function (1.2). When $r(\|X\|^2) \equiv 1$, this is just $\delta_{a,C,D}(X)$ given in the previous section by (3.1.1). If $r(\|X\|^2)$ is a non-decreasing function between 0 and 1 (not including 0), and $r(\|X\|^2)/\|X\|^2$ is a non-increasing function and in addition, conditions (1) and (2) of Theorem 3.1.1 are satisfied then $\delta_{a,r,C,D}(X)$ is a better estimator of θ than the minimax estimator X with respect to general quadratic loss (1.2). This will be restated and proven in the following theorem.

Theorem 3.2.1: If $X = [X_1, X_2, \ldots, X_p]$ has a s.s. distribution about θ and the loss is general quadratic loss (1.2), then the estimator $\delta_{a,r,C,D}(X)$ given by (3.2.1) is a better estimator of θ than X for $p \geq 4$, provided:

(1) $0 < r(\cdot) \leq 1$

(2) $r(\|X\|^2)$ is non-decreasing

(3) $r(\|X\|^2)/\|X\|^2$ is non-increasing

(4) $\operatorname{tr} CD > 2\xi_L = 2$(maximum eigenvalue of $D^{\frac{1}{2}}CD^{\frac{1}{2}}$)

and (5) $0 < a \leq ((2/p)(\operatorname{tr} CD-2\xi_L)\gamma_L^{-1})/E_0(\|X\|^{-2})$

where, as in Theorem 3.1.1, γ_L is the maximum eigenvalue of $D^{\frac{1}{2}}C^2D^{\frac{1}{2}}$.

<u>**Proof:**</u> It follows, by similar techniques to those used in the proof of Theorem 3.1.1 that if the $p \times 1$ random vector $X \sim$ s.s. about θ, then

$$((RD(X,\ \theta)-RD(\delta_{a,r,C,D},\ \theta))((p-1)/a)$$

$$= 2\sum_{i=1}^{p} \xi_i \theta_i^{*2}\|\theta\|^{-2}E[r(((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)R^2)(p-1)(Z_1^2+\|\theta\|R^{-1}Z_1)((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)^{-1}]$$

$$+2\sum_{i=1}^{p} \xi_i(1-\theta_i^{*2}\|\theta\|^{-2})E[r(((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)R^2)\|T\|^2((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)^{-1}]$$

(3.2.2)
$$-a[\sum_{i=1}^{p}\gamma_i \theta_i^{**2}\|\theta\|^{-2}E[R^{-2}r(((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)R^2)(p-1)(Z_1+\|\theta\|R^{-1})^2((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)^{-2}]$$

$$+\sum_{i=1}^{p}\gamma_i(1-\theta_i^{**2}\|\theta\|^{-2})E[R^{-2}r(((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)R^2)\|T\|^2((Z_1+\|\theta\|R^{-1})^2+\|T\|^2)^{-2}]]$$

where, (1) $\theta^* = [\theta_1^*, \theta_2^*, \ldots, \theta_p^*]' = P'\theta$ and P is the $p \times p$ orthogonal matrix such that $P'D^{\frac{1}{2}}CD^{\frac{1}{2}}P$ is the diagonal matrix whose terms along the diagonal are $\xi_1, \xi_2, \ldots, \xi_p$, the eigenvalues of $D^{\frac{1}{2}}CD^{\frac{1}{2}}$; (2) $\theta^{**} = [\theta_1^{**}, \theta_2^{**}, \ldots, \theta_p^{**}]' = Q'\theta$ and Q is the $p \times p$ orthogonal matrix such that $Q'D^{\frac{1}{2}}C^2D^{\frac{1}{2}}Q$ is the diagonal matrix whose terms along the diagonal are $\gamma_1, \gamma_2, \ldots, \gamma_p$ the eigenvalues of $D^{\frac{1}{2}}C^2D^{\frac{1}{2}}$; (3) $Z = [Z_1, Z_2, \ldots, Z_p]' \sim \mathcal{U}\{\|Z\|^2 = 1\}$ and $\|T\|^2 = \sum_{i=2}^{p} Z_i^2$; and (4) $R = \|X-\theta\|$.

Define $S(Z_1, \theta, R)$ by

$$S(Z_1,\ \theta,\ R) = (Z_1+\|\theta\|R^{-1})^2 + \|T\|^2 = 1+2\|\theta\|R^{-1}Z_1 + \|\theta\|^2R^{-2}.$$ By this definition and since $\sum_{i=1}^{p}(\theta_i^{**})^2 = \|\theta\|^2$ then

$$\sum_{i=1}^{p}\gamma_i E[R^{-2}r(R^2 s(Z_1,\ \theta,\ R))\|T\|^2(1-(\theta_i^{**})^2\|\theta\|^{-2})(s(Z_1,\ \theta,\ R))^{-2}]$$

$$\leq (p-1)\gamma_L E[R^{-2}r(R^2 s(Z_1,\ \theta,\ R))\|T\|^2(s(Z_1,\ \theta,\ R))^{-2}].$$

It follows directly by this inequality and (3.2.2), that for $0 < a \leq (2/p)(\text{tr } CD-2\xi_L)\gamma_L^{-1}/E_0(\|X\|^{-2}) = (2/p)(\text{tr } CD-2\xi_L)\gamma_L^{-1}/E(R^{-2})$ that

$$(RD(X,\ \theta)-RD(\delta_{a,r,C,D},\ \theta))((p-1)/a)$$

(3.2.3)
$$\geq 2\sum_{i=1}^{p}\xi_i(\theta_i^{*})^2\|\theta\|^{-2}E[r(R^2 s(Z_1,\ \theta,\ R))((p-1)(Z_1^2+\|\theta\|R^{-1}Z_1)-\|T\|^2)(s(Z_1,\theta,R))^{-1}]$$
$$+2\ \text{tr}(CD)E[r(R^2 s(Z_1,\ \theta,\ R))\|T\|^2(s(Z_1,\ \theta,\ R))^{-1}]$$
$$-(2/p)(\text{tr}(CD)-2\xi_L)(p-1)(E(R^{-2}))^{-1}E[R^{-2}r(R^2 s(Z_1,\ \theta,\ R))(s(Z_1,\ \theta,\ R))^{-1}]$$

Moreover, it was proven in the proof of Theorem 2.2.1 that

(3.2.4)
$$(E(R^{-2}))^{-1}E[R^{-2}r(R^2 s(Z_1,\ \theta,\ R))(s(Z_1,\ \theta,\ R))^{-1}]$$
$$= (E(R^{-2}))^{-1}E[r(\|X\|^2)\|X\|^{-2}]$$
$$\leq E[R^2 r(\|X\|^2)\|X\|^{-2}]$$
$$= E[r(R^2 s(Z_1,\ \theta,\ R))(s(Z_1,\ \theta,\ R)^{-1})]$$

Combining (3.2.3) and (3.2.4),

$$(RD(X,\ \theta)-RD(\delta_{a,r,C,D},\ \theta))((p-1)/a)$$

(3.2.5)
$$\geq 2\sum_{i=1}^{p}\xi_i\ (\theta_i^{*})^2\|\theta\|^{-2}E[r(R^2 s(Z_1,\theta,R))(pZ_1^2+(p-1)\|\theta\|R^{-1}Z_1-1)(s(Z_1,\theta,R))^{-1}]$$
$$+\ 2\ \text{tr}(CD)E[r(R^2 s(Z_1,\theta,R))(1-Z_1^2)(s(Z_1,\theta,R))^{-1}]$$
$$-(2/p)(\text{tr}(CD)-2\xi_L)(p-1)E[r(R^2 s(Z_1,\theta,R)(s(Z_1,\theta,R))^{-1}]$$

Note that by property (3) of this theorem,

$$
\begin{aligned}
0 &\geq (d/dZ_1)r(R^2 S(Z_1,\theta,R))(S(Z_1,\theta,R))^{-1} \\
(3.2.6)\qquad &= 2\|\theta\|Rr'(R^2 S(Z_1,\theta,R))(S(Z_1,\theta,R))^{-1} - 2\|\theta\|R^{-1}r(R^2 S(Z_1,\theta,R))(S(Z_1,\theta,R))^{-2}
\end{aligned}
$$

In addition, using the density of Z_1 given in Lemma A.4, and integrating by parts, the following is obtained:

$$
\begin{aligned}
&E[r(R^2 S(Z_1,\theta,R))(pZ_1^2-1)(S(Z_1,\theta,R))^{-1}|R] \\
(3.2.7)\qquad &= M^* \int_{-1}^{1} (1-z_1^2)^{\frac{p-1}{2}}\, 2\|\theta\| z_1 [Rr'(R^2 S(z_1,\theta,R))(S(z_1,\theta,R))^{-1} \\
&\quad - R^{-1} r(R^2 S(z_1,\theta,R))(S(z_1,\theta,R))^{-2}]dz_1
\end{aligned}
$$

and

$$
\begin{aligned}
&(p-1)E[r(R^2 S(Z_1,\theta,R))Z_1(S(Z_1,\theta,R))^{-1}|R] \\
(3.2.8)\qquad &= M^* \int_{-1}^{1} (1-z_1^2)^{\frac{p-1}{2}}\, 2\|\theta\| [Rr'(R^2 S(z_1,\theta,R))(S(z_1,\theta,R))^{-1} \\
&\quad - R^{-1} r(R^2 S(z_1,\theta,R))(S(z_1,\theta,R))^{-2}]dz_1
\end{aligned}
$$

where $M^* = [\int_0^1 (1-z_1^2)^{\frac{p-3}{2}} dz_1]^{-1}/2$.

From (3.2.6), (3.2.7), and (3.2.8), clearly,

$$(3.2.9)\qquad E[r(R^2 S(Z_1,\theta,R))Z_1(S(Z_1,\theta,R))^{-1}] \leq 0$$

and

$$(3.2.10)\qquad E[r(R^2 S(Z_1,\theta,R))(pZ_1^2+(p-1)Z_1-1)(S(Z_1,\theta,R))^{-1}] \leq 0.$$

Using these facts, the difference in risks will now be shown to be non-negative for two cases.

<u>Case 1:</u> $E[r(R^2S(Z_1,\theta,R))(pZ_1^2+(p-1)\|\theta\|R^{-1}Z_1-1)(S(Z_1,\theta,R))^{-1}] \leq 0$.

By (3.2.9) and (3.2.10), when $\|\theta\| \geq R$ this is always true. Since

$\sum_{i=1}^{p} \xi_i(\theta_i^{**})^2\|\theta\|^{-2}$ is less than or equal to ξ_L, from (3.2.5), letting

$(\mathrm{tr}\ CD / \xi_L) = q$,

$$(RD(X, \theta)-RD(\delta_{a,r,C,D}, \theta))((p-1)/a)$$

$$\geq (2\xi_L/p)[pE[r(R^2S(Z_1,\theta,R))(pZ_1^2+(p-1)\|\theta\|R^{-1}Z_1-1)(S(Z_1,\theta,R))^{-1}]$$

$$+pqE[r(R^2S(Z_1,\theta,R))](1-Z_1^2)(S(Z_1,\theta,R))^{-1}]$$

$$-(q-2)(p-1)E[r(R^2S(Z_1,\theta,R))(S(Z_1,\theta,R))^{-1}]]$$

$$= (2\xi_L/p)Dq.$$

If $E[r(R^2S(Z_1,\theta,R))(1-pZ_1^2)S(Z_1,\theta,R)]$ is non-negative then, since property (4) implies $q > 2$,

$$Dq \geq D_2 = pE[r(R^2S(Z_1,\theta,R))((p-2)Z_1^2+(p-1)\|\theta\|R^{-1}Z_1+1]$$

$$\geq p(p-1)E[r(R^2S(Z_1,\theta,R))(2Z_1^2+\|\theta\|R^{-1}Z)(S(Z_1,\theta,R))^{-1}]$$

$$= p(p-1)R\|\theta\|^{-1}E[r(R^2S(Z_1, \theta,R))[Z_1-Z_1(S(Z_1, \theta, R))^{-1}]]$$

$$\geq 0.$$

The last inequality is due to (3.2.9) and property (2), which leads to

$$E[r(R^2S(Z_1,\theta,R))Z_1|R] \geq E[r(R^2S(Z_1,\theta,R))|R]E[Z_1|R] = 0.$$

If $E[r(R^2S(Z_1,\theta,R))(1-pZ_1^2)(S(Z_1,\theta,R))^{-1}]$ is negative,

$$Dq \geq D_p = (p-1)E[r(R^2S(Z_1,\theta,R))(p\|\theta\|R^{-1}Z_1+2)(S(Z_1,R,\theta))^{-1}]$$

$$= (p/2)(p-1)[\text{Difference in risks between } X \text{ and } \delta_{a,r}(X)$$

$$\text{when the loss is quadratic loss and } a = (2(p-2)/p)R^2]$$

$$\geq 0.$$

The last term was shown to be non-negative in the proof of Theorem 2.2.1.

Case 2: $E[r(R^2S(Z_1,\theta,R))(pZ_1^2+(p-1)\|\theta\|R^{-1}Z_1-1)(S(Z_1,\theta,R)^{-1}] \geq 0$.

This only occurs when $\|\theta\| \leq R$ and by (3.2.9) and (3.2.10),

$$\sum_{i=1}^{p} \xi_i(\theta_i^*)^2\|\theta\|^{-2}E[r(R^2S(Z_1,\theta,R))(pZ_1^2+(p-1)\|\theta\|R^{-1}Z_1-1)(S(Z_1,\theta,R))^{-1}]$$

$$\geq \xi_L E[r(R^2S(Z_1,\theta,R))(pZ_1^2+(p-1)Z_1-1)(S(Z_1,\theta,R))^{-1}].$$

For this case, $E[r(R^2S(Z_1,\theta,R))(1-pZ_1^2)(S(Z_1,\theta,R))^{-1}]$ must be negative and so,

$$(R(X,\theta)-R(\delta_{a,r,C,D},\theta))((p-1)/a)$$

$$(3.2.11) \quad \geq (2\xi_L/p)(p-1)E[r(R^2S(Z_1,\theta,R))(pZ_1+2)(S(Z_1,\theta,R))^{-1}]$$

$$\geq (2\xi_L/p)(p-1)E[r(R^2S((-2/p),\theta,R))\ E[(pZ_1+2)(S(Z_1,\theta,R))^{-1}|R]].$$

Again, using the density of Z_1 given in Lemma A.4,

$$E[(pZ_1+2)(S(Z_1,\theta,R))^{-1}|R]$$

$$(3.2.12) \quad = M^*\int_{-1}^{1}(1-z_1^2)^{\frac{p-3}{2}}(pz_1+2)(1+2\|\theta\|_1z_1+\|\theta\|_1^2)^{-1}dz_1$$

$$= M^*\int_{0}^{1}(1-z_1^2)^{\frac{p-3}{2}}(4(1+\|\theta\|_1^2)-4p\|\theta\|_1z_1^2)((1+\|\theta\|_1)^2-4\|\theta\|_1^2z_1^2)^{-1}dz_1$$

where $M^* = [\int_0^1(1-z_1^2)^{\frac{p-3}{2}}dz_1]^{-1}/2$ and $\|\theta\|_1 = \|\theta\|R^{-1}$.

If, as in (2.1.6), $d(\|\theta\|_1, z_1) = (1+\|\theta\|_1)^2-4\|\theta\|_1^2z_1^2$, then by Lemma A.3, the random variable Z_1, having the density

$$g_{p-2,1}(z_1) = \begin{cases} ((1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1,z_1))/\int_0^1((1-z_1^2)^{\frac{p-3}{2}}/d(\|\theta\|_1,z_1))dz_1 & \text{when } 0 \leq z_1 \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

has MLR non-decreasing in Z_1 when $\|\theta\|_1 \leq 1$.

So, $E_{\|\theta\|_1}Z_1^2 \leq E_1 Z_1^2 = 1/(p-2)$, where the expectation is taken with respect to the density $g_{p-2,1}(z_1)$. From (3.2.12),

$$E[(pZ_1+2)(S(Z_1,\theta,R))^{-1}|R]$$

$$\propto 4((1+\|\theta\|_1^2) - p\|\theta\|_1 E_{\|\theta\|_1} Z_1^2)$$

$$\geq (4/(p-2))((p-2)-p\|\theta\|_1+(p-2)\|\theta\|_1^2)$$

$$\geq (8/(p-2))(1-\|\theta\|_1)^2$$

$$\geq 0.$$

Applying this to (3.2.11), for this case,

$$(R(X,\theta)-R(\delta_{a,r,C,D},\theta))((p-1)/a) \geq 0.$$

In addition, from (3.2.11), when $\theta = 0$,

$$(R(X,0)-R(\delta_{a,r,C,D},0))/((p-1)/a)$$

$$= (2\xi_L/p)(p-1)E[r(R^2)(pZ_1+2)]$$

$$= (4\xi_L/p)(p-1)Er(R^2) > 0.$$

Thus, $\delta_{a,r,C,D}(X)$ is better than X, and the proof of this theorem is complete.

4. Multiple observations. For one $p \times 1$ $(p \geq 4)$ observation X on a spherically symmetric distribution about θ, minimax estimators which are better than X with respect to quadratic (1.1) and general quadratic (1.2) loss were found in sections 2 and 3. In this section, the problem of estimating the mean of a p-dimensional $(p \geq 4)$ spherically symmetric distribution when n observations are taken will be considered and reduced to a one-observation problem.

In practice, one would usually sample n observations $X_1, X_2, \ldots, X_n$ from a s.s. distribution about θ, and estimate θ using some estimator which depends on all these observations. For this case, the best invariant procedure is Pitman's estimator given by, $\delta(X_1, X_2, \ldots, X_n) = X_1 - E_0[X_1 | Y_2, Y_3, \ldots, Y_n]$ where $Y_i = X_i - X_1$, $i=2,3, \ldots, n$. However, this best invariant procedure is often very difficult to calculate and other estimators such as the sample mean $\bar{X}$ or a maximum likelihood estimate (MLE) may be preferred. All these estimators just mentioned belong to a larger class of estimators which have s.s. distributions about θ. Specifically, these estimators are contained in the class C of spherically symmetric translation invariant estimators. Thus, $\delta(X_1, \ldots, X_n) \in C$ if

(4.2.1) $\delta(PX_1, PX_2, \ldots, PX_n) = P\delta(X_1, X_2, \ldots, X_n)$ where P is a $p \times p$ orthogonal matrix

and

(4.2.2) $\delta(X_1 - C, X_2 - C, \ldots, X_n - C) = \delta(X_1, X_2, \ldots, X_n) - C$ where C is a $p \times 1$ vector.

If $\delta = \delta(X_1, X_2, \ldots, X_n) \in C$, then δ will have a s.s. distribution about θ. This will easily be proven in the following theorem.

Theorem 4.1: If $X_1, X_2, \ldots, X_n$ are n i.i.d. random vectors having a p-dimensional spherically symmetric distribution about θ, and $\delta(X_1, X_2, \ldots, X_n)$ is an estimator of θ satisfying (4.2.1) and (4.2.2) then $\delta(X_1, X_2, \ldots, X_n)$ also has a spherically symmetric distribution about θ.

Proof: By Definition 2.1, $\delta(X_1, \ldots, X_n)$ has a s.s. distribution about θ if $P(\delta(X_1, X_2, \ldots, X_n)-\theta)$ has the same distribution as $\delta(X_1, X_2, \ldots, X_n)-\theta$ for any fixed $p \times p$ orthogonal matrix P. By assumption, $\delta(X_1, X_2, \ldots, X_n)$ satisfies (4.2.1) and (4.2.2), thus for any set S,

$$\Pr(P(\delta(X_1, X_2, X_n)-\theta) \in S)$$

$$= \Pr(\delta(P(X_1-\theta), P(X_2-\theta), \ldots, P(X_n-\theta)) \in S)$$

$$= \Pr(\delta(X_1-\theta), \delta(X_2-\theta), \ldots, (X_n-\theta)) \in S)$$

$$= \Pr((\delta(X_1, X_2, \ldots, X_n)-\theta) \in S).$$

The last two equalities follow by (4.2.2) and the fact that $P(X_i-\theta)$ has the same distribution as $(X_i-\theta)$. Thus, $\delta(X_1, X_2, \ldots, X_n)$ has a s.s. distribution about θ. Q. E. D.

Pitman's estimator and $\overline{X}$ clearly satisfy (4.2.1) and (4.2.2). If the MLE is unique, it is immediate from the definition of an MLE, that it too satisfies (4.2.1) and (4.2.2). Thus all these estimators have s.s. distributions about θ.

It is not clear that there will always be a spherically symmetric translation invariant MLE. However, one does exist and this will now be proven.

Theorem 4.2: If $X_1, X_2, \ldots, X_n$ is a random sample from a p-dimensional s.s. distribution about θ, then if there exists at least one MLE there exists an MLE having a s.s. distribution about θ.

Proof: By definition, $\hat{\theta}(X_1, X_2, \ldots, X_n)$ is an MLE of θ if

$$\max_{\theta} \prod_{i=1}^{n} f(X_i - \theta) = \prod_{i=1}^{n} f(X_i - \hat{\theta}(X_1, X_2, \ldots, X_n))$$

where X_i has the density $f(\cdot)$. If the MLE is unique, it is clear that it satisfies (4.2.1) and (4.2.2) and thus has a s.s. distribution about θ.

Suppose there exists more than one MLE. Let $\hat{\theta}_I(X_1, X_2, \ldots, X_n)$ be an MLE. For every orthogonal matrix P there exists an MLE $\hat{\theta}_P$ such that

$$\hat{\theta}_I(PX_1, PX_2, \ldots, PX_n) = P\hat{\theta}_P(X_1, X_2, \ldots, X_n).$$

If $X_1, X_2, \ldots, X_n$ is some given point, then

$$O(X_1, X_2, \ldots, X_n) = \left\{ (y_1, y_2, \ldots, y_n) : (y_1, y_2, \ldots, y_n) = (PX_1, PX_2, \ldots, PX_n) \text{ for some orthogonal matrix } P \right\}$$

$$= \text{orbit mapped out by } (X_1, X_2, \ldots, X_n).$$

Consider a set R_I, containing exactly one point from each orbit. Thus define some property and let $R_I = \{(X_1, X_2, \ldots, X_n) : (X_1, X_2, \ldots, X_n) \text{ satisfies property I}\}$ where if property I is satisfied by at least one $X_1, X_2, \ldots, X_n$ from each orbit, then there does not exist an orthogonal matrix $P \neq I$ such that $X_1, X_2, \ldots, X_n \neq PX_1, PX_2, \ldots, PX_n$ and $PX_1, PX_2, \ldots, PX_n$ satisfies property I. Clearly, one can define such a property. (For example, $X_1, X_2, \ldots, X_n$ satisfies property I if $X_i = [X_{i1}, X_{i2}, \ldots, X_{ip}]$ and for any Z_n on the same orbit, $X_{i1} > Z_{i1}$ and $X_1, X_2, \ldots, X_n$ is unique. If the uniqueness is not satisfied add the condition $X_{i2} > Z_{i2}$ and keep adding such conditions until property I is satisfied by a unique $X_1, X_2, \ldots, X_n$). For any $p \times p$ orthogonal matrix define

$$R_P = \left\{ (y_1, y_2, \ldots, y_n) : (y_1, y_2, \ldots, y_n) = (PX_1, PX_2, \ldots, PX_n) \text{ for some } (X_1, X_2, \ldots, X_n) \in R_I \right\}$$

Let $\hat{\theta}(X_1, X_2, \ldots, X_n) = \hat{\theta}_{P'}(X_1, X_2, \ldots, X_n)$ when $(X_1, X_2, \ldots, X_n) \in R_P$. Clearly, $\hat{\theta}(X_1, X_2, \ldots, X_n)$ is an MLE and property (4.2.1) is satisfied (as it is satisfied for all MLE's). Moreover, if Q is any $p \times p$ orthogonal matrix and for some P, $QX_1, QX_2, \ldots, QX_n \in R_P$ then

$$\hat{\theta}(QX_1, QX_2, \ldots, QX_n) = \hat{\theta}_{P'}(QX_1, QX_2, \ldots, QX_n)$$

$$= P\hat{\theta}_I(P'QX_1, P'QX_2, \ldots, P'QX_n) = Q\hat{\theta}_{P'Q}(X_1, X_2, \ldots, X_n)$$

$$= Q\hat{\theta}(X_1, X_2, \ldots, X_n) \text{ since } (X_1, X_2, \ldots, X_n) \in R_{Q'P}.$$

Thus, $\hat{\theta}(X_1, X_2, \ldots, X_n)$ is a s.s. translation invariant MLE, and so by theorem 4.1, it has a s.s. distribution about θ. Q.E.D.

Hence, when sampling from a s.s. distribution about θ, all spherically symmetric translation invariant estimators have s.s. distributions and included in this class, are many of the most commonly used estimators, such as $\bar{X}$, Pitman's estimator and an MLE. Estimators which are better than these estimators in 4 or more dimensions may be obtained by applying the results of sections 2 and 3. For example, if the loss is quadratic loss (1.1) and $\delta = \delta(X_1, X_2, \ldots, X_n)$ has a spherically symmetric distribution about θ, the James-Stein estimator

$$\delta_a(\delta) = (1-a||\delta||^{-2})\delta$$

is a better estimator of θ than δ for $0 < a \leq (2(p-2)/p)E_0||\delta||^{-2}$. This is a direct application of Theorem 2.1.1. Note too, that if $E_0||\delta||^{-2}$ is not known, but, a bound can be placed on it (say $E_0||\delta||^{-2} \leq b_\delta$), then $\delta_a(\delta)$ is better than δ for $0 < a \leq (2(p-2)/p)/b_\delta$. In fact, this class of estimators improves on all s.s. translation invariant estimators δ, satisfying $E_0||\delta||^{-2} \leq b_\delta$.

5. Remarks. In section 4, the multiple observation case was reduced to one observation by showing that if $\delta(X_1, X_2, \ldots, X_n)$ is any spherically symmetric translation invariant estimator, it has a s.s. distribution about θ. Thus, the discussion in this section will be restricted to one observation X from a $p \times 1$ s.s. distribution.

Note that the bounds on the class of minimax estimators presented in sections 2 and 3 are the best bounds for the general s.s. case. The estimator $\delta_{a,r,C,D}(X) = (I - ar(||X||^2)D^{-\frac{1}{2}}CD^{\frac{1}{2}}/||X||^2)X$, as given by (3.2.1), was proven to be better than X with respect to general quadratic loss (1.2) provided conditions 1-5 of Theorem 3.2.1 are satisfied. The bound on this class was given by Condition 5 which states that $0 < a \leq ((2/p)(\mathrm{tr}\ CD - 2\xi_L)\gamma_L^{-1})/E_0(||X||^{-2})$. Bock's result for the multivariate normal distribution (given in [8]), requires $0 < a \leq ((2/(p-2))(\mathrm{tr}\ CD - 2\xi_L)\gamma_L^{-1})/E_0(||X||^{-2})$, hence, our class of estimators differs from the normal class of estimators by a factor of $(p-2)/p$. However, the best estimator when sampling from a normal population occurs when $a = (p-2)^{-1}(\mathrm{tr}\ CD - 2\xi_L)\gamma_L^{-1}/E_0(||X||^{-2})$ which is included in the general class when $p \geq 4$. In addition, by an appropriate choice of the matrix of C (suggested by James Berger), $\delta_{a,r,C,D}(X)$ may be reduced to a simpler estimator which will have no requirements on D except that it be positive definite. Specifically, allow $C = D^{-1}/d_L$ where d_L is the maximum eigenvalue of D^{-1}, then $\xi_L = \gamma_L = d_L^{-1}$ and $\delta_{a,r,D^{-1}/d_L,D} = (I - (ar(||X||^2)D^{-1})/(d_L||X||^2))X$. With this choice of C, Condition 4 of Theorem 3.2.1 is satisfied $(\mathrm{tr}\ CD = p/d_L > 2\xi_L = 2/d_L)$ and provided Conditions (1)-(3) hold, this estimator is better than X for $0 < a \leq (2(p-2)/p)/E_0||X||^{-2})$. So, a simple estimator exists for general quadratic loss (1.2).

A direct application of the results of sections 2 and 3 is linear regression. That is, these minimax estimators may be used to improve on the least squares estimator of the regression coefficient when the normal assumptions of the linear model may not be true. Consider the linear regression model $Y = X\beta + \epsilon$ where Y is $n \times 1$, X is $n \times p$, β is $p \times 1$ and ϵ is $n \times 1$ and has a s.s. distribution about 0. The least squares estimator of β is $\hat{\beta} = (X'X)^{-1}X'Y$ which has a p-dimensional s.s. distribution about β. Thus, for $p \geq 4$, a number of estimators which are better than $\hat{\beta}$ are available. For example, $\delta_{a,r}(\hat{\beta}) = (1-ar(\|\hat{\beta}\|^2)\|\hat{\beta}\|^{-2})\hat{\beta}$ has a smaller risk than $\hat{\beta}$ with respect to quadratic loss (1.1), when $p \geq 4$ and the conditions of Theorem 2.2.1 are satisfied.

The importance of the results of this paper is twofold. When taking one observation from a s.s. distribution about θ, there exist simple estimators of the mean which are better than the usual procedure, and, in addition, the multiple observation case so simply reduces to one observation and improvements can be made over many of the most commonly used estimators based on a number of observations. Moreover, due to the publication of various papers, such as [14] by Efron and Morris, which explicitly show the practical value of the James-Stein estimator, these improved estimators are becoming more familiar and widely used.

A. Appendix. In this section,, lemmas containing important properties which aid in the proof of Theorems in sections 2 and 3 are given. Some of these lemmas are special cases of ones given in the Appendix of [12].

Lemma A.1. Suppose $X = [X_1, X_2, \ldots, X_p]$ has a uniform distribution over the sphere $(\|X\|^2 \leq 1)$, then the joint density of X_1 and $Z = \|X\|^2$ is given by

$$(A.1)\qquad f(x_1, z) = \begin{cases} (M/2)(z-x_1^2)^{\frac{p-3}{2}} & \text{when } -\sqrt{z} < x_1 \leq \sqrt{z} \text{ and } 0 \leq z \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

where $M = [(2/p)\int_0^1 (1-y^2)^{\frac{p-3}{2}} dy]^{-1}$.

Proof: This density is given in Lemma 6.1.6 in [12]. This is the special case when $\|\theta\| = 0$ and $R = 1$.

Lemma A.2. When $p \geq 3$,

$$\int_0^1 ((1-y^2)^{\frac{p-3}{2}} / ((1-\|\theta\|^2)^2 + 4\|\theta\|^2(1-y^2)))dy$$

$$= \begin{cases} (\int_0^1 (1-y^2)^{\frac{p-3}{2}} dy)[h(\|\theta\|, 1)]p & \text{when } \|\theta\| > 1 \\ (\int_0^1 (1-y^2)^{\frac{p-3}{2}} dy)[h(1, \|\theta\|)]p & \text{when } \|\theta\| \leq 1 \end{cases}$$

where

$$[h(\|\theta\|, 1)]p = (\|\theta\|^2(1+\|\theta\|^2))^{-1} \sum_{i=0}^{\infty} (-1)^i a_i \|\theta\|^{-2i}$$

and

$$[h(1, \|\theta\|)]p = (1+\|\theta\|^2)^{-1} \sum_{i=0}^{\infty} (-1)^i a_i \|\theta\|^2$$

and $a_0 = 1$ and $a_i = [(p-2(i+1))/(p+2(i-1))]a_{i-1}$ for $i=1,2, \ldots$.

Proof: This is just a restatement of Lemma 6.1.5 in [12] when $R = 1$.

Lemma A.3. If Y is a random variable with a density with respect to Lebesque measure given by

$$g_{(2q+1,r)}(y) = \begin{cases} ((1-y^2)^q/d(\|\theta\|,y)^r)/\int_0^1((1-y^2)^q/(d(\|\theta\|,y))^r)dy & \text{when } 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

where, $d(\|\theta\|,y) = (1-\|\theta\|^2)^2+4\|\theta\|^2(1-y^2)$, then the distribution of Y has monotone likelihood ratio (MLR) non-decreasing in Y when $\|\theta\| \le 1$ and MLR non-increasing in Y when $\|\theta\| > 1$.

Proof: The proof of this lemma is straightforward.

Lemma A.4. If the $p \times 1$ random vector $X = [X_1, X_2, \ldots, X_p]'$ has a uniform distribution over the entire area of the sphere $(\|X\|^2 \le 1)$, then the density of $Z_1 = X_1/\|X\|$ is given by

$$\text{(A.2)} \qquad g(z_1) = \begin{cases} M^*(1-z_1^2)^{\frac{p-3}{2}} & \text{when } -1 \le z_1 \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

where $M^* = [\int_0^1 (1-y^2)^{\frac{p-3}{2}} dy]^{-1}/2$.

Proof: By Lemma A.1, if $Z_1 = \|X\|^2$, then the joint density of X_1 and Z as given by (A.1) is

$$f(x_1, z) = \begin{cases} (M/2)(z-x_1^2)^{\frac{p-3}{2}} & \text{for } -\sqrt{z} \le x_1 \le \sqrt{z} \text{ and } 0 \le z \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

where $M = [(2/p)\int_0^1 (1-y^2)^{\frac{p-3}{2}} dy]^{-1}$. If a transformation to $Z_1 = X_1/\sqrt{Z}$ and $Z = Z$ is made, the joint density of Z_1 and Z is

$$f(z_1, z) = \begin{cases} (M/2)z^{\frac{p-2}{2}}(1-z_1^2)^{\frac{p-3}{2}} & \text{for } -1 \leq z_1 \leq 1 \text{ and } 0 \leq z \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Thus, when $-1 \leq z_1 \leq 1$, the density of z_1 is

$$g(z_1) = ((M/2)\int_0^1 z^{\frac{p-2}{2}}\, dz)(1-z_1^2)^{\frac{p-3}{2}} = M^*(1-z_1^2)^{\frac{p-3}{2}}. \qquad \text{Q.E.D.}$$

<u>Lemma A.5.</u> If $X = [X_1, X_2, \ldots, X_p]' \sim \mathcal{U}\{\|X\|^2 \leq 1\}$ and $Z_i = X_i/\|X\|$ for $i = 1,2,\ldots,p$, then for any integrable function $r(\cdot)$

$$(p-1)E[r(\|Z+\theta\|^2)(Z_i^2+\theta_i Z_i)\|Z+\theta\|^{-2}]$$

$$\text{(A.3)} \quad = (\theta_i^2/\|\theta\|^2)E[r((Z_1+\|\theta\|)^2+\|T\|^2)(p-1)(Z_1^2+\|\theta\|Z_1-\|T\|^2)((Z_1+\|\theta\|)^2+\|T\|^2)^{-1}]$$

$$+E[r((Z_1+\|\theta\|)^2+\|T\|^2)((Z_1+\|\theta\|)^2+\|T\|^2)^{-1}]$$

and

$$(p-1)E[r(\|Z+\theta\|^2)(Z_i+\theta_i)^2\|Z+\theta\|^{-4}]$$

$$\text{(A.4)} \quad = (\theta_i^2/\|\theta\|^2)E[r((Z_1+\|\theta\|)^2+\|T\|^2)((p-1)(Z_1+\|\theta\|)^2-\|T\|^2)((Z_1+\|\theta\|)^2+\|T\|^2)^{-2}]$$

$$+E[r((Z_1+\|\theta\|)^2+\|T\|^2)\|T\|^2((Z_1+\|\theta\|)^2+\|T\|^2)^{-2}]$$

where $Z = [Z_1, Z_2, \ldots, Z_p]'$, $\theta = [\theta_1, \theta_2, \ldots, \theta_p]'$ and $\|T\|^2 = \sum_{i=2}^{p} Z_i^2$.

<u>Proof:</u> Expressions (A.3) and (A.4) will be gotten by making two transformations of variables. First, let P be a $(p-1) \times (p-1)$ orthogonal matrix such that $P[\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p]' = [\|\theta\|_i, 0, \ldots 0]'$ where $\|\theta\|_i = \sqrt{\|\theta\|^2-\theta_i^2}$, and transform to $S = [S_2, S_3, \ldots, S_p]' = P[Z_1,Z_2,\ldots,Z_{i-1},Z_{i+1},\ldots,Z_p]'$. Second, there exists a $p \times p$ orthogonal matrix Q, such that $Q[\theta_1,\|\theta\|_i, 0, \ldots 0]' = [\|\theta\|, 0, \ldots 0]'$, and if $Y = [Y_1, Y_2, \ldots, Y_p]' = Q[Z_i, S_2, S_3, \ldots, S_p]'$ then $Z_i = (\theta_i/\|\theta\|)Y_1-(\|\theta\|_i/\|\theta\|)Y_2$, $S_2 = (\|\theta\|_i/\|\theta\|)Y_1+(\theta_i/\|\theta\|)Y_2$ and $S_i = Y_i$ for $i = 3, 4, \ldots, p$. The fact that Y and Z have the same distribution and

$$E[r((Z_1+\|\theta\|)^2+\|T\|^2)Z_1Z_2((Z_1+\|\theta\|)^2+\|T\|^2)^{-q}] = 0 \quad \text{for} \quad q = 1,2$$

and $$E[r((Z_1+\|\theta\|)^2+\|T\|^2)Z_2((Z_1+\|\theta\|)^2+\|T\|^2)^{-q}] = 0 \quad \text{for} \quad q = 1,2,$$

gives the desired results. Q.E.D.

Lemma A.6. If $X = [X_1, X_2, \ldots, X_p]' \sim \mathcal{U}\{\|X\|^2 \leq 1\}$, $Z_i = X_i/\|X\|$, $\|T\|^2 = \sum_{i=2}^{p} Z_i^2 = 1-Z_1^2$, and $\|\theta\| = (\sum_{i=1}^{p} \theta_i^2)^{\frac{1}{2}}$, then

$$E[((p-1)(Z_1^2+\|\theta\|Z_1)-\|T\|^2)((Z_1+\|\theta\|)^2+\|T\|^2)^{-1}] \leq 0.$$

Proof: Rewriting the expectation with $\|T\|^2 = 1-Z_1^2$ and using the density for Z_1 given in Lemma A.4,

$$E[(p-1)(Z_1^2+\|\theta\|Z_1)-\|T\|^2]$$
$$\propto \int_0^1 (1-z_1)^{\frac{p-3}{2}} [((p-1)(1-\|\theta\|^2)-(p-(p-2)\|\theta\|^2)(1-z_1^2))/d(\|\theta\|, z_1)]dz_1$$

where $d(\|\theta\|, z_1) = (1+\|\theta\|^2)^2-4\|\theta\|^2 z_1^2$. It is clear from Lemma 6.1.8 proven in [12] and the above integral expression that

$$E[(p-1)(Z_1^2+\|\theta\|Z_1)-\|T\|^2]$$

$$\propto -E_0[(p-1)(S_1+\|\theta\|)^2-\|Y\|^2]$$

$$\leq 0$$

where $S = [S_1, S_2, \ldots, S_p]' \sim \mathcal{U}\{\|S\|^2 \leq 1\}$ and $\|Y\|^2 = \sum_{i=2}^{p} S_i^2$. Q.E.D.

Lemma A.7. If $X = [X_1, X_2, \ldots, X_p]' \sim \mathcal{U}\{\|X\|^2 \leq 1\}$ and $Z_i = X_i/\|X\|$, then

$$E[((p-1)(Z_1+\|\theta\|)^2-\|T\|^2)((Z_1+\|\theta\|)^2+\|T\|^2)^{-2}] \geq 0.$$

Proof: The proof of this follows directly by the MLR properties of $g_{p-2,1}(Z_1)$ given in Lemma A.3.

References

[1] Alam, K. (1973). "A Family of Minimax Estimators of the Mean of a Multivariate Normal Distribution," Ann. Statist. 1, 517-525.

[2] Baranchik, A.J. (1964). "Multiple Regression and Estimation of the Mean of a Multivariate Normal Distribution," Stanford University, Technical Report No. 51.

[3] Baranchik, A.J. (1964). "A Family of Minimax Estimators of the Mean of a Multivariate Normal Distribution," Ann. Math. Statist. 41, 642-645.

[4] Berger, J. (1974). "Minimax Estimation of a Multivariate Normal Mean with Arbitrary Quadratic Loss," Technical Report No. 376, Purdue University. (Submitted to the Annals of Statistics.)

[5] Berger, J. (1975). "Minimax Estimation of Location Vectors for a Wide Class of Densities," Ann. Statist. 3, 1318-1328.

[6] Berger, J. (1976). "Admissible Minimax Estimation of a Multivariate Normal Mean with Arbitrary Quadratic Loss," Ann. Statist. 4, 223-226.

[7] Bhattacharya, P.K. (1966). "Estimating the Mean of a Multivariate Normal Population with General Quadratic Loss Function," Ann. Math. Statist. 37, 1819-1824.

[8] Bock, M.E. (1973). "Minimax Admissible Estimators of the Mean of a Multivariate Normal Distribution with Non-Identity Covariance Matrix," University of Illinois.

[9] Bock, M.E. (1975). "Minimax Estimators of the Mean of a Multivariate Normal Distribution," Ann. Statist. 3, 209-218.

[10] Brieman, L. (1968). Probability. Addison-Wesley Publishing Company, Inc.

[11] Brown, L.D. (1966). "On the Admissibility of Invariant Estimators of One or More Location Parameters," Ann. Math. Statist. 37, 1087-1136.

[12] Cohen, A.R. and Strawderman, W.E. (1975). "Minimax Estimation of Location Parameters for Spherically Symmetric Unimodal Distributions," Technical Report No. 259, University of Minnesota. (Submitted to the Annals of Statistics.)

[13] Dempster, A.P. (1969). Elements of Continuous Multivariate Analysis. Addison-Wesley Publishing Company, Inc.

[14] Efron, B. and Morris, C. (1975)."Data Analysis Using Stein's Estimator and Its Generalizations," Journal of the Amer. Statistical Assoc., 70, 311-319.

[15] James, W. and Stein, C. (1961). "Estimation with Quadratic Loss," Proc. Fourth Berkeley Symp. of Math. Stat. and Probability, Vol. 1, University of California Press, Berkely and Los Angeles, 361-379.

[16] Kiefer, J. (1957). "Invariance, Minimax Sequential Estimation, and Continuous Time Processes," Ann. Math. Statist. 28, 537-601.

[17] Lehmann, E.L. (1959). Testing Statistical Hypotheses. New York: John Wiley and Sons, Inc.

[18] Stein, C. (1955). "Inadmissibility of the Usual Estimator for the Mean Vector of a Multivariate Normal Distribution," Proc. Third Berkeley Symp. on Math. Stat. and Probability, Vol. 1., University of California Press, Berkeley and Los Angeles, 197-206.

[19] Strawderman, W.E. (1971). "Proper Bayes Minimax Estimators of the Multivariate Normal Mean," Ann. Math. Statist. 42, 385-388.

[20] Strawderman, W.E. (1974). "Minimax Estimation of Location Parameters for Certain Spherically Symmetric Distributions," Journal of Multivariate Analysis, 1-10.